PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/00 — A2

(11) International Publication Number: WO 00/64928

(43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/GB00/01546

(22) International Filing Date: 20 April 2000 (20.04.00)

(30) Priority Data:
9909161.3 — 21 April 1999 (21.04.99) — GB

(71) Applicants (*for all designated States except US*): SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB). MAX-PLANCK INSTITUTE OF NEUROBIOLOGY [DE/DE]; Am Klopferspitz 18A, D-82152 Planegg-Martinsried (DE).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BARDE, Yves-Alain [CH/DE]; Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, D-82152 Planegg-Martinsried (DE). FRIEDEL, Roland [DE/DE]; Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, D-82152 Planegg-Martinsried (DE). MICHALOVICH, David [GB/GB]; SmithKline Beecham Pharmaceuticals, New Frontiers Science Park South, Third Avenue, Harlow, Essex CM19 5AW (GB). REITH, Alastair, David [GB/GB]; SmithKline Beecham Pharmaceuticals, New Frontiers Science Park South, Third Avenue, Harlow, Essex CM19 5AW (GB). SCHNUERCH, Harald [DE/DE]; Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, D-82152 Planegg-Martinsried (DE). STUBBUSCH, Jutta [DE/DE]; Max-Planck Institute of Neurobiology, Am Klopferspitz 18A, D-82152 Planegg-Martinsried (DE).

(74) Agent: CONNELL, Anthony, Christopher; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL COMPOUNDS

(57) Abstract

R35 polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing R35 polypeptides and polynucleotides in therapy, and diagnostic assays for such.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

N vel Compounds

Field f the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers, e.g., cAMP (Lefkowitz, Nature, 1991, 351:353-354). Herein these proteins are referred to as proteins participating in pathways with G-proteins or PPG proteins. Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and dopamine (Kobilka, B.K., et al., Proc. Natl Acad. Sci., USA, 1987, 84:46-50; Kobilka, B.K., et al., Science, 1987, 238:650-656; Bunzow, J.R., et al., Nature, 1988, 336:783-787), G-proteins themselves, effector proteins, e.g., phospholipase C, adenyl cyclase, and phosphodiesterase, and actuator proteins, e.g., protein kinase A and protein kinase C (Simon, M.I., et al., Science, 1991, 252:802-8).
For example, in one form of signal transduction, the effect of hormone binding is activation of the enzyme, adenylate cyclase, inside the cell. Enzyme activation by hormones is dependent on the presence of the nucleotide GTP. GTP also influences hormone binding. A G-protein connects the hormone receptor to adenylate cyclase. G-protein was shown to exchange GTP for bound GDP when activated by a hormone receptor. The GTP-carrying form then binds to activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself, returns the G-protein to its basal, inactive form. Thus, the G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector, and as a clock that controls the duration of the signal.

The membrane protein gene superfamily of G-protein coupled receptors has been characterized as having seven putative transmembrane domains. The domains are believed to represent transmembrane α-helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

G-protein coupled receptors (otherwise known as 7TM receptors) have been characterized as including these seven conserved hydrophobic stretches of about 20 to 30 amino acids, connecting at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological

disorders. Other examples of members of this family include, but are not limited to, calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins, endothelial differentiation gene-1, rhodopsins, odorant, and cytomegalovirus receptors.

Most G-protein coupled receptors have single conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed to stabilize functional protein structure. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 has been implicated in signal transduction.

Phosphorylation and lipidation (palmitylation or farnesylation) of cysteine residues can influence signal transduction of some G-protein coupled receptors. Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several G-protein coupled receptors, such as the b-adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

For some receptors, the ligand binding sites of G-protein coupled receptors are believed to comprise hydrophilic sockets formed by several G-protein coupled receptor transmembrane domains, said socket being surrounded by hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is postulated to face inward and form polar ligand binding site. TM3 has been implicated in several G-protein coupled receptors as having a ligand binding site, such as the TM3 aspartate residue. TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson et al., Endoc. Rev., 1989, 10:317-331) Different G-protein a-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors have been identified as an important mechanism for the regulation of G-protein coupling of some G-protein coupled receptors. G-protein coupled receptors are found in numerous sites within a mammalian host.

Over the past 15 years, nearly 350 therapeutic agents targeting 7 transmembrane (7 TM) receptors have been successfully introduced onto the market.

There is a continuing need to identify and characterise further 7TM receptor genes and their related polypeptides/proteins, as targets for drug discovery.

Summary f the Invention

The present invention relates to R35, a novel 7TM receptor, in particular R35 polypeptides and R35 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of neuropathic pain, inflammatory and chronic pain, incontinence, disorders of the gastrointestinal tract associated with gut motility and secretion (for example irritable bowel syndrome) hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with R35 imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate R35 activity or levels. In a yet further aspect the invention relates to the use of the polynucleotides and polypeptides of the invention to identify agonists or antagonists for use in therapy.

Description of the Invention

In a first aspect, the present invention relates to R35 polypeptides. Such peptides include isolated polypeptides comprising an amino acid sequence which has at least 95% identity, preferably at least 97-99% identity, to that of SEQ ID NO:2, 4, 6 or 11 over the entire length of SEQ ID NO:2, 4, 6 or 11 respectively. Such polypeptides include those comprising the amino acid sequence of SEQ ID NO:2, 4, 6 or 11.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 95% identity, preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2, 4, 6 or 11 over the entire length of SEQ ID NO:2, 4, 6 or 11 respectively. Such polypeptides include the polypeptides of SEQ ID NO:2, 4, 6 or 11.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1, 3, 5 or 10.

Polypeptides of the present invention are believed to be members of the family of polypeptides. The ligand binding and signalling properties of R35 are hereinafter referred to as "R35 activity" or "R35 polypeptide activity" or "biological activity of R35". Also included amongst the activities of R35 are antigenic and immunogenic activities of said R35 polypeptides, in particular the antigenic and immunogenic activities of the polypeptide of SEQ ID NO:2, 4, 6 or 11. Preferably, a polypeptide of the present invention exhibits at least one biological activity of R35.

The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a fusion protein. It is often advantageous to include an

additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The present invention also includes include variants of the aforementioned polypetides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to R35 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 95% identity to the amino acid sequence of SEQ ID NO:2, 4, 6 or 11 over the entire length of SEQ ID NO:2, 4, 6 or 11 respectively. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1, 3, 5or 10 encoding the polypeptides of SEQ ID NO:2, 4, 6 or 11 respectively.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 95% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, 4, 6 or 11, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 95% identity to SEQ ID NO:1, 3, 5 or 10 over the entire length of SEQ ID NO:1, 3, 5 or 10 respectively. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such

polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1, 3, 5 or 10 as well as the polynucleotides of SEQ ID NO:1, 3, 5 or 10.

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequences are as follows:

SEQ ID NO:1 is a chick cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 510 to 2678) encoding a polypeptide of 723 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2.

SEQ ID NO:3 is a mouse cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1 to 2196) encoding a polypeptide of 732 amino acids, the polypeptide of SEQ ID NO:4. The nucleotide sequence encoding the polypeptide of SEQ ID NO:4 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:3 or it may be a sequence other than the one contained in SEQ ID NO:3, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:4.

SEQ ID NO:5 is a rat cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1 to 2190) encoding a polypeptide of 730 amino acids, the polypeptide of SEQ ID NO:6. The nucleotide sequence encoding the polypeptide of SEQ ID NO:6 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:5 or it may be a sequence other than the one contained in SEQ ID NO:5, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:6.

SEQ ID NO:10 is a partial human cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 1 to 1824) encoding a polypeptide of 608 amino acids, the polypeptide of SEQ ID NO11:. The nucleotide sequence encoding the polypeptide of SEQ ID NO:11 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:10 or it may be a sequence other than the one contained in SEQ ID NO:10, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:11.

The nucleotide sequence of SEQ ID NO:7 is a partial chick cDNA sequence isolated using RNA fingerprinting methodology.

The nucleotide sequence of SEQ ID NO:8 is a partial mouse cDNA sequence encoding a polypeptide (SEQ ID NO:9) that is homologous to chick R35.

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one R35 activity.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(1989)), from a cDNA library derived from embryonic day 16 chick dorsal root ganglia (DRG). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2, 4, 6 or 11 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, 3, 5 or 10 may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention Typically these nucleotide sequences are greater than 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides.

A polynucleotide encoding a polypeptide of the present invention may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization

conditions with a labeled probe having the sequence of SEQ ID NO: 1, 3, 5, 10 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1, 3, 5, 10 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is cut short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™ technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise

a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al, Basic Methods in Molecular Biology (1986) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (supra). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the

cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:10 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeledR35 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton*et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprisingR35 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility tothe Diseases through detection of mutation in the R35 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagonostic kit which comprises:
(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 10, or a fragment thereof ;
(b) a nucleotide sequence complementary to that of (a);
(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:11 or a fragment thereof; or
(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:11.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly neuropathic pain, inflammatory and chronic pain, incontinence, disorders of the gastrointestinal tract associated with gut motility and secretion (for example, irritable bowel syndrome), amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Th differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor*et al.*, Immunology Today (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to

polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention are responsible for one or more biological functions, including one or more disease states, in particular the Diseases hereinbefore mentioned. It is therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide of SEQ ID NO:2, 4 or 6 or a human homolog, for example a polypeptide comprising the polypeptide of SEQ ID NO:11. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned.

Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan*et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polpypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring R35 activity in the mixture, and comparing the R35 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and R35 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide

sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypetide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2, 4, 6 or 11.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an interative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, neuropathic pain, inflammatory and chronic pain, incontinence, disorders of

the gastrointestinal tract associated with gut motility and secretion (for example, irritable bowel syndrome), related to either an excess of, or an under-expression of, R35 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the R35 polypeptide.

In still another approach, expression of the gene encoding endogenous R35 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or separately administered (see, for example, O'Connor, J Neurochem (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices with the gene can be supplied (see, for example, Lee *et al.*, Nucleic Acids Res (1979) 6:3073; Cooney *et al.*, Science (1988) 241:456; Dervan *et al.*, Science (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*.

For treating abnormal conditions related to an under-expression ofR35 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activatesa polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of R35 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read,

BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the

sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such as GCG and Lasergene. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1, 3, 5 or 10 and/or a polypeptide sequence encoded therefrom.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or

oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, Proteins - Structure and Molecular Properties, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in Post-translational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", Meth Enzymol (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", Ann NY Acad Sci (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the

sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., SIAM *J. Applied Math., 48:* 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research 12(1):* 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol. 215:* 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol. 215*: 403-410 (1990). The well known Smith Waterman algorithm may also be used to determine identity.

Preferred parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Preferred parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, J. Mol Biol. 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

By way of example, a polynucleotide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:1, that is be 100% identical, or it may include up to a certain integer number of nucleotide alterations as compared to the reference sequence. Such alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleotide alterations is determined by multiplying the total number of nucleotides in SEQ ID NO:1 by the numerical percent of the respective percent identity(divided by 100) and subtracting that product from said total number of nucleotides in SEQ ID NO:1, or:

$$\mathbf{n}_n \leq \mathbf{x}_n - (\mathbf{x}_n \bullet \mathbf{y}),$$

wherein $\mathbf{n}_n$ is the number of nucleotide alterations, $\mathbf{x}_n$ is the total number of nucleotides in SEQ ID NO:1, and $\mathbf{y}$ is, for instance, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%,etc., and wherein any non-integer product of $\mathbf{x}_n$ and $\mathbf{y}$ is rounded down to the nearest integer prior to subtracting it from $\mathbf{x}_n$. Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

Similarly, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2, that is be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the % identity is less than 100%. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or

insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 by the numerical percent of the respective percent identity(divided by 100) and then subtracting that product from said total number of amino acids in SEQ ID NO:2, or:

$$\mathbf{n}_a \leq \mathbf{x}_a - (\mathbf{x}_a \bullet \mathbf{y}),$$

wherein $\mathbf{n}_a$ is the number of amino acid alterations, $\mathbf{x}_a$ is the total number of amino acids in SEQ ID NO:2, and $\mathbf{y}$ is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and wherein any non-integer product of $\mathbf{x}_a$ and $\mathbf{y}$ is rounded down to the nearest integer prior to subtracting it from $\mathbf{x}_a$.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a subject sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the sequences being compared as hereinbefore described. Falling within this generic term are the terms "ortholog", meaning a polynucleotide or polypeptide that is the functional equivalent of a polynucleotide or polypeptide in another species, and "paralog" meaning a functionally similar sequence when considered within the same species.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

Examples

Example 1 - Cloning of partial chick R35 cDNA by RNA fingerprinting

The initial R35 fingerprint fragment was identified in an RNA-fingerprinting approach, comparing RNA from NGF and NT3 treated DRG explants, as described (Friedel, R.H., Schnurch, H., Stubbusch, J., and Barde, Y.-A., Proc. Natl. Acad. Sci. USA, 1992; 94: 12670-12675).

Experimental design

DRG Explant Cultures

Lumbosacral ganglia 4, 5, and 6 of E10 chicken embryos were carefully dissected, washed in PBS, and transferred to 48-well tissue culture plates (Falcon) coated with polyornithine (Sigma) and mouse laminin (Life Technologies) according to Ernsberger, U. & Rohrer, H. (1988) Dev. Biol. 126, 420-432.

Each well contained 0.5 ml of culture medium consisting of Ham's F-14 (Life Technologies) supplemented with 0.1 mg/ml transferrin, 16 ug/ml putrescine, 6 ng/ml progesterone, 8 ng/ml sodium selenite, 100 ug/ml penicillin G, 60 ug/ml streptomycin sulfate (all reagents from Sigma), and 2 ng/ml NGF or NT3. After 4 days of incubation at 37°C in a 3% CO2 atmosphere at 95% relative humidity, ganglia were detached from the wells with a small spatula and total RNA was isolated by using an RNeasy kit (Qiagen).

RNA Fingerprinting.

cDNAs were synthesized from total RNA with oligo(dT) primers and Moloney murine leukemia virus reverse transcriptase (Superscript II; Life Technologies) at 45°C for 1 hour. Reaction products were purified from excess oligonucleotides by using a QIAquick kit from Qiagen. cDNA derived from 60 ng of total RNA was used as template for a PCR. Amplification was performed in a volume of 25 ul containing 1x PCR buffer (Perkin-Elmer), 20 uM dNTPs (Pharmacia), 50 nM each primer, and 5 uCi (1 uCi = 37 kBq) of [α-33P]dATP (Amersham). For each reaction, two 18-mer oligonucleotides of arbitrary sequence and a G+C content of 61% were used (see list below). All PCRs were run on thermal cyclers 2400/9600 from Perkin-Elmer. After 3 minutes of denaturation at 95°C, 1.5 units of AmpliTaq DNA polymerase (Perkin-Elmer) was added (manual "hot start"). After 2 minutes at 95°C the profile was started with two low-stringency cycles (95°C for 30 seconds, 50°C for 2 minutes, and 72°C for 2 minutes) followed

by 25-30 high-stringency cycles (95°C for 30 seconds, 65°C for 1 minute, and 72°C for 1 minute), and a final 5-minute extension at 72°C. Reaction products were separated by electrophoresis on nondenaturing 4% polyacrylamide gels for 3 hours at 40 W. Gels were vacuum dried on chromatography paper (3MM; Whatman), fixed in a cassette, and overlaid with x-ray film. Exposure times varied from overnight to 3 days. Bands of interest were marked on the gel by piercing through the developed film. Gel pieces were excised with a scalpel and DNA was eluted by boiling in 50 ul of 10 mM TrisHCl, 1 mM EDTA buffer, pH 7.5, for 5 minutes. One microliter of the eluted DNA was used as template in a second PCR (reamplification). The reaction volume was 50 ul containing 1x PCR buffer, 0.2 mM dNTPs, and 0.1 uM primers (the same as used for first PCR). Then 1.5 units of AmpliTaq were added by manual "hot start." The temperature profile consisted of 30 cycles (95°C for 15 seconds, 65°C for 30 seconds, and 72°C for 30 seconds) and a final 5-minute extension at 72°C. A 10ul aliquot of the reaction was separated on an ethidium bromide-stained 1% agarose gel. Occasionally, reamplification products contained by-products of lower molecular weight. These were excluded by cutting out only the band of the highest molecular weight. The DNA was purified with a QIAquick kit and cloned in the pCRII vector of the TA-Cloning kit (Invitrogen). DNA sequencing was done with an automated sequencer (ABI 373A; Applied Biosystems). Primers used for RNA fingerprinting:

P1 5' AGTCGCAGCACAGGTGAG;
P2 5' CGTCCGTAGGTCCAGGT;
P3 5' CCCAGGTCGTCGTGTTCA;
P4 5' GTCAGGGATACCGAGAGC;
P5 5' CCAGCCGCCAGAGTTTTG;
P6 5' GAGAGTTAGCCGCAGCG;
P7 5' GTTACAGCAACCGCCAGG;
P8 5' GGGGTCCTTCAATGGGAG;
P9 5' TCCGAGGGAGATTCCGTC; and
P10 5' GCCACAGGGCAGAGTCTT.

The initial 248 nucleotide cDNA fragment of chick R35 (SEQ ID NO:7) was cloned by this means from the RNA of NGF-treated DRG explants of E10 chicken embryos with primer pair P3 and P7 (NB: the primer sequences are not included in the sequence of SEQ ID NO:7; P3 is the 5' primer and P7 is the 3' primer).

Example 2: Cloning of full length (ie. including entire coding region) chick R35 cDNA

The fingerprint fragment isolated from Example 1 comprising the 248 nucleotide chick R35 cDNA sequence (SEQ ID NO:7) served as a probe to screen a lambda gt10 cDNA-library from chicken E16 DRGs. Subsequently 3 cDNA clones were obtained, that overlap and cover 4.5 kb of the R35 cDNA. Two of these lambda clones (called R35lambda-13 and R35lambda-12) cover the entire open reading frame. The screening of the lambda library was essentially done as described (Sambrook. *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). The full-length chick R35 cDNA sequence is given in SEQ ID NO:1.

Example 3: Cloning of partial mouse R35 cDNA

The 431 nucleotide mouse partial cDNA fragment (SEQ ID NO:8) was cloned by PCR, using mouse genomic DNA as template. Non-degenerate primers were deduced from conserved regions in the alignment of the full length chicken R35 cDNA seqeunce with the sequences of other species we have cloned (i.e. goose, nuthatch, ostrich, snake and turtle). The primers used for the PCR were:

SK-R35-S2; 5' TTTGCCTCTCTGTGGTTGGG;

N-R35-A3; 5' GAGCCAAAGAATGACTTTGGTCAG.

PCR conditions:

300 ng mouse genomic DNA was used as template. The reaction volume was 50 ul containing 1x PCR buffer(Perkin Elmer) with 2.5mM Mg2+, 0.2 mM dNTPs, and 0.2 uM of each primer SK-R35-S2 and N-R35-A3. 1.5 units of AmpliTaq were added by manual "hot start." The temperature profile consisted of 40 cycles (95°C for 30 seconds, 63°C for 30 seconds, and 72°C for 1 minute).

Example 4: Cloning of full-length (ie. the entire coding region) mouse R35 cDNA

Full length mouse R35 cDNA was cloned by PCR extension of the initial fragment from example 3 with using a lamda gt10 cDNA library from mouse P0 brain (constructed by H. Schnuerch) as template. For this purpose R35 specific primers from the fragment obtained in example 3 were designed and used in combination with lambda gt10 specific primers which flanked the cDNA insertion site of the lambda DNA.

Each cDNA-PCR extension consisted of two rounds of PCR: a first PCR with an aliquot of the lamda gt10 cDNA library as template, and a R35 specific primer and a lambda specific primer. The second PCR was done with an aliquot of the first PCR as template and nested primers (again one R35 specific and one lambda specific).

PCR conditions were cycles of 95°C 15s, 65°C 30s, and 72°C 1 min.
PCR products were seperated by gel electrophoresis in an agarose gel, the band with the highest molecular weight was cut out and cloned in the pCRII vector (Invitrogen).

Primers for the mouse R35 5' extension:
1. PCR: L1 and MM-R35-EU1
2. PCR: L2 and MM-R35-EU2
The resulting PCR fragment was designated "EU"

Primers for the R35 3' extension:
1. PCR: L1 and MM-R35-ED1
2. PCR: L2 and MM-R35-ED2
The resulting PCR fragment was designated "ED"

The full-length mouse R35 cDNA sequence was compiled from EU + fragment obtained in example 3 + ED.
The sequence was confirmed by RT-PCR with mouse R35 specific primers from an independent cDNA of P3 mouse brain.

Primer sequences:
The lambda specific primers:
L1: AAG AGC TGA CGC AAG TTC TGG TA (for first PCR)
L2: TTG AGC AAG TTC AGC CTG GTT AAG (nested, for second PCR)

The mouse R35 specific primers for 5' extension:
MM-R35-EU1: TATACCGCGAAGAGGAGCAGCA (for first PCR)
MM-R35-EU2: AAGGCGTGCGCACGAAGACA (nested, for second PCR)

The mouse R35 specific primers for 3' extension:
MM-R35-ED1: GCACCAACAGCTTCCCCTTG (for first PCR)
MM-R35-ED2: GAGCTTGGCACAGAAGCGCT (nested, for second PCR).

Example 5: Cloning of rat R35 cDNA

The rat R35 cDNA was cloned by RT-PCR with a rat P2 cDNA library, using mouse primers under low stringency conditions: cycles of 95°C 15s, 56°C 30s, and 72°C 1 min.
The mouse primers used to amplify the rat cDNA:
MM-R35-S6: TCAGACTCGTTGAAGAGGGA (upstream primer)
MM-R35-A15: CAAAAACTCATGCACGAGAAC (downstream primer)

The sequences of MM-R35-S6 and -A15 are not included in the rat R35 cDNA sequence given in SEQ ID NO5; they are located in the 5' and 3' untranslated regions respectively.

Example 6: Cloning of human HS-R35 cDNA
An initial human R35 cDNA fragment of 223 nt was cloned by RT-PCR with primers from the murine R35 sequnce, using cDNA of adult human brain as template.

Mouse R35 primers used to clone the initial human fragment:
MM-R35-S2: TGGGAATGGTGCAGGAGTAA
MM-R35-A3: ACACTGGTGTCCGCAAACGT

PCR conditions were standard: cycles of 95°C 15s, 65°C 30s, and 72°C 1 min
The 5' extension of the initial cDNA was done by RT-PCR using a adult human brain cDNA library, cloned in the expression vector pcDNA3.1 (Invitrogen), as template. Primers were human specific (HS-R35-A1) and vector specific (PCT7).

Primer sequences:
HS-R35-A1: ACACAAGGGAATGGCAAGGG
PCT7: TAATACGACTCACTATAGGGAGACC

PCR conditions were standard: cycles of 95°C 15s, 65°C 30s, and 72°C 1 min.

The sequence obtained was a large fragment of the human sequence, encompassing the 5' coding region, encoding the N-terminal 608 amino acid residues. The polynucleotide sequence of this fragment is shown in SEQ ID NO:10 and the polypeptide encoded thereby is shown in SEQ ID NO:11.

The 3' human cDNA sequence, which together with the sequence disclosed in SEQ ID NO:10 will provide the complete full-length human sequence, may be cloned according to methods well known in the art. For example a 3' RACE reaction may be performed using a PCR primer sequence specific to the human sequence in combination with a vector specific primer from a suitable library. From this information a single full-length clone may be obtained, for example by PCR, or the 5' and 3' fragments may be joined using methods well known in the art.

Example 7: Expression profile of (chick) R35 mRNA by in situ hybridisation

Preparation of tissue sections and hybridisation with single stranded RNA-probes were performed as described (Schnurch, H. and Risau, W., 1991; Development 111: 1143-1154). The probe, R35-S1-A1, used for in situ hybridisation was synthesised from a 478 nucleotide cDNA fragment of the R35 open reading frame (nucleotides 1123 to 1600 in SEQ ID NO:1).

Results:

Probing transverse sections of chick embryonic dorsal root ganglia (DRGs) at days 6, 8, 10, 12, 14 and 16 of gestation revealed R35 mRNA to be restricted to a specific subpopulation of DRG neurons. At all developmental stages examined, this R35-positive population was concentrated in the medial part of the DRG. The medial neurons (also called mediodorsal neurons) are those that express the NGF receptor trkA and are equivalent to the "small" NGF-dependent neurons of rodents. This subpopulation is generally assumed to be responsible for noci- and thermoception. R35 may therefore have a role in noci- or thermoception. No R35 mRNA was detectable in the lateral part of the DRG.

Transverse sections of chick embryonic spinal cord at 6 and 8 days gestation revealed robust R35 expression detectable in the entire spinal cord, including the meninges. In 10 day gestation embryos, a relatively weak R35-specific hybridization signal was detected throughout the spinal cord, but expression remained high in the motoneuron pool. No R35 signal was detectable in the meninges at this stage. At 12, 14 and 16 days gestation, R35 expression was restricted to the motoneuron pool with no signal detectable in other regions of spinal cord, or in the meninges. The marked decrease in R35 expression from E8 to E10 in the spinal cord is concommitant with the end of proliferation and the onset of terminal differentiation in this region. This fact would suggest a potential role of R35 in the regulation of proliferation/differentiation of spinal cord neurons.

In the developing brain at 10 days gestation, R35 expression was detected in the tectum where it was restricted to two layers of neurons, which lie close to the ventricle. Other regions of

the tectum showed no detectable R35 expression. At this stage of development R35 mRNA was also detected in retina, being mainly restricted to the retinal ganglion cell layer. Other layers in the retina showed only very weak, or no, signal. At 14 days gestation a strong R35-specific hybridisation signal was detected in the external granule cell layer of the cerebellum.. No other regions of the cerebellum show R35 signal at this stage. The external granule cell layer consists mainly of proliferating granule cells, that subsequently will move inside the cerebellum to form the mature granule cell layer.

R35 mRNA was also detected in neurons of the developing gut. At 12 days gestation the R35 hybridization signal in the gut was restricted to two concentric layers of neurons, which presumably correspond to the plexus myentericus (outer ring) and the plexus submucosus (inner ring). Within developing gonads, R35 hybridization signal was evenly distributed in the gonad anlage at 6 days gestation and expression remained detectable in cells of the gonads at 12 days gestation.

Conclusions: these data validate findings from initial RNA fingerprinting and confirm R35 as a novel seven transmembrane receptor. Restricted expression of R35 in peripheral, central and autonomic nervous systems implicate functions in molecular processes and signalling pathways that regulate these neuronal tissues. As such, modulation of R35 function may be of value for therapeutic intervention in relation to neuropathic pain, inflammatory and chronic pain, incontinence, disorders of the gastrointestinal tract associated with gut motility and secretion (for example, irritable bowel syndrome).

SEQUENCE INFORMATION

SEQ ID NO:1

CTAAGAGGCAGCTCTGTCTGGGAAGATACTCACTCCCCATTTCTCTGCTATCCTGCTTGC
TTCTGACTCCTCCCTCCTGCAGCTCAGATCAGCACAGGGTGAATCCAACAGCCTCCCATT
TCTACCGTGTTTCTCCCTGAGCTTGAGCATACCTGGTACTCTGCCTTCGTTTCAGCTGAA
CTCGTGAGTAAGGCACATTGCCTGATATCTGTCAAAAGTGAATGGCTTGTTATTACCTGA
AGAGACTGAATTAACTATCTTCTGGTACTGAAAGGGTAATGAAGAACTGTATTTTTTCCA
GCACCTTTTGTACTTCTTTGTTGCCTGATTTTTCAAGAGCGTGAGTTATTCTGGTTATCA
AGGGAGCTTAAACCTGAGTATTTCTTCACCGGATGCTACAAAAGAAGAAAGGAAAACAAA
CTTAAGATGTGGAGCTAGCAGTTATGGTTTCAGCTTGTGAAAACTTGCAGAGTCAGTGAG
AAGGCAAGGAACAACATTTGCCAAGGACCATGTCGGTAATGCCTAGTAATTTGTCCCTCA
ATGGGACAAGCTTCTTCGCTGAAAATCATAGCATTATGGATAAACCTAATGAGCAGAGAA
CTTTGAATGTCTTTCTATTTTGCTCGACTTTTATCATTGCATTCACAGTGCTTCTGGGCA
GCATTTATTCACTAGTTTCCCTGCTGAAACTGCAGAACAAATCCACAATTTCAATGATTG
TGACCTCTTTGTCAATAGATGATTTGATCAGCATTGTGCCAGTGATTATTTTCATGCTCA
CCCAGTGGTCAAGTGATGCCCTTCCTCAGCCTCTGTGCACCACCTCAGCACTTATATATT
TATTCCAGGGCATTTCTAGCAATCTGAAAGGGTCTCTTATAGTTTCCTACAACTTCTATA
GCATCAACAAAACTGAGACAATGAACTGCAGTGCTTCCAAACGACGAGTGAGCATGGTAT
GGGCAATTCTTAGCATTTGGATCGTCAGTTTGCTGATATGCATTTTGCCTCTCTGTGGCT
GGGGCAAATACATCCCCACCACCTGGGGTTGCTTCACAGACCATGCAAGTTCCTACATCT
TGTTCTTATTTATTGTCTACTCATTGTGCTTTTGCCTCCTCACAGTGCTGTCTGTCCCTC
TAACTTACCAGCTGTTGTGCTCAGATGAGCAACAGCTATTACACGTCAACTACCAGGAAA
TCTCTCGAGGCTACATCACCCCTGGGACACCTGCAGGCTGCAGTACTGCCACCCCATGTC
TGTCACCAGTGGACCCTGTGGATAAAACCCTGAAGCATTTCCAAAATGCATGTCTAAGCT
CAAAGGCAGTTTTTAGGAAAGGTGTGGCTGAGAGCAGTGGACTCGAGCCCCGATGCATGA
ACAACGCAAAGAGCAGGAGCTTCACAGTGGGCTTTGCCCAGAAACGGTTCTCACTGATTC
TTGCATTGACAAAAGTCATTCTCTGGCTCCCAATGATGATACAAATGGTTGTCCAGCACA
TCACTGGCTATCAAAGCTTTTCATTTGAGACACTTAGCTTCCTGCTGACATTGCTAGCTG
CCACCGTCACCCCAGTATTTGTCCTGTCAGAGCACTGGATCCACCTGCCGTGTGGCTGCA
TCATAAACTGCAGGAGGAATTCATATGCTGTGTCTTCAGAAGAACTCAAAACCAAGCATA
GAGGTTTCGAATTCAACCTTTCCTTCCAGCATGGTTATGGAATCTACCGAATATCCCCCG
AGAGCCACCACCACGATGGCGATGGTAAATCCACATCGTGCCACAACCTGCTGGTTTGCG
AGAAACCGTACGAGCCGCCGAGGGGAGGTGGCGGCGGTGCGGCGCTGGCGGAACTCAGCA
CCACGGATAGCGCCCGGCCAGGCCCGGCGGGGCTCAGCGCCACGGACAGCGCCAGGCCCG
GTCCCGCGGGGTTCAGCACCACAGACAGCGCCAGGCCCAGCGCCGCGGGGGTGCGCGGGG
AGGCGGCCTCGGGCCTCGGCAGGACCGTGGAGGGACCCGAGAGGAGGCTGTCGCACGAAG
AGGGCCATAAGCCAGAGCTCACAGACTGGGAGTGGTGCAGGAGTAAATCAGAGAGGACCC
CTCGACAGCGGTCAGGTGGCGCCTTAGCTATTCCTCTGTGTGCATTTCAAGGAACCGTGT
CACTTCAGGCACCCACAGGAAAAACTCTCTCCTTGTCTACATACGAGGTAAGCACTGAAG
GACAGAAAATAACACCCACATCAAAGAAAATTGAAGTTTATCGATCTAAAAGTGTTGGTC
ATGACCCCAACCCAGAGGAGTGTCCTAATACGTTTGCTGACACGAGTGTTAAAATTCATT
TAGAGGTTCTGGAAATTTGTGACAATGAAGAGGCTCTGGACACTGTTTCAATCATCAGTA
ATATCAGCCAGTCCTCCACACAGGTAAGATCTCCATCCTTACGCTACTCACGGAAGGAGA

ACAGATTTGTCTCATGTGATTTGGGGGAAACTGCTTCCTACTCACTCTTCATACCATCCA
ACAATCCCGACAGCGATATTAACATAACAATTCCAGACACTGTGGAAGCTCACCGGCAGA
ATAGTAAAAAGCAACACATGGAAAGAGGTGGTTATCAGGAAGAAATACAGATGTTGAATA
AAGCGTACAGAAAACGGGAAGAAGATGGCAACAGCAATTAAAAGACATTTAATCTAAACA
TTTAAACTGATTCAGCAGACATTGCTCCTCTTGACTCAAGTTACCCTAACAAATACAATG
CCAACCGGTTTGTTTTTTTTTTTTTCCTAGCATATAAATGATTTTTATTTGGGTACACTG
TTAAATAGTTAACATTTACAAATGGGAAAACAAACGCTTCAGTGCTTTGCAGCTTTGTGC
AATGAAATCCAAGTTATTTACCAATCTCTTTATTTGAGGAAGACTTGCATCTGAAAGGAA
TCCTAAATGAACTCTTAAAAGCCACCCTTGCTTGCTTGTTGGCACTGGGTGGTCTGTGGG
AAGCCAGAAAATGAGAGAGCAGCTCTACACCCCTGAAGATGTAACAGGTTTCAATGACTG
TGGCAGTTCTCCTACTGATGTGCCAGACTTGAAACATTTACTGAAAATATGTAGCACAAG
CCTAAGACTAAGCCATGCATTTGGTTAGGGATTGTAAGTTACCTACGTGGTTTAATTTTC
TACTGCTCCATACGTAATTGATACGTAATGTCTCTGTAAATGGCCATTGTATAGACCTTT
GAATGTGGGTGTCTGAAATTTAAGCTTGCACATTTTATATGTCCAGCACTGTGAAAATGT
GTTCTCATGGCTGGACATTTCTTTCATCGAGTGTGTCAAATGTTGGTTGATCTTATGCTG
GTTACTCTAAAGTGTGCCATTTTCTGCCTCCTTGATCTCTTAGGAGAGGGGACAGGGCTG
CTCAAGGAGGCCAAACTTAGAGACCCAGCTTAGCTGTTGACCTCAGCTTGAGGCTGAGTG
CCCAGATCAGCTTCCCTGCCCTGGAGCACCTCTGGAGGAGCCTGTTCTTATGCTGCCTTA
CAGGAAGCGTAGAGCCAAGTCTGAAATCAAACTATAACCCACCAATAGACTGCTTCTTGT
CAAATTCAAACAGACTGAGGGTAAAACAGCAAATGAAACACTTAAGTCTAATGCTTTTTG
TCTGAGAGCTCATTTTCTGAGATGTTGTCATACATTTCAGCCAGATAGAAGTAGACTTCA
GCAGGACTTTAAGAGGTTTGACTAAAAGTTACACCACTTCTTTAATTAATAGATTTAATT
TTCAGTGAGTGACAACATAATAATTCAGGTGTCTGCAGTCACCTTGTTATCCATTCAGTG
GGAACATGCACAGGAAAGTAAGCACAGCCCTATTTAGGATTCCTTTATGGATTGTGAACA
GGTAGAGCAAAAGGGCCAACTAGCTCAGCTTGTGTTCAGTATGGAGGCATGGAAAGAACA
GGTGATTTGCATTCTCACCTCACCAACATAAATCAGGCCTGCCACCATGATCAGACGGGC
TATGCTGCTGTGACAGTGATGGAGAGCTGAAGAGAATCAGGGTCAGAGACTGCAGGCCGC
TGTGATGGAGATGGATCACTCTGTGCAAGATGTTGCTTTTCACGTGCTGCTCTTCCAAGG
TTGAAAACAAAGTTCCTCCATAAGATTGCCTCAAGGCTCCTTAGTACCTGGTGATGCTGC
GTTACACTTCAGATTTATCAAGAAACTATAGTAGACAGAGCAATCAATGGTAAGTGTATT
AGTAGGTTATATCCATGTGCATGCAGCAGGGAGACACGTTCTGCTGTCTCTGTATTACTT
TGCTTTGTGTTTGTAGGATTAGCTCAGAAAGTTGCTTTATAGCAACAGATGTGCTCAATC
CAGAAAGACGTTGTGAGGGAGGATGCAGTGCATCACTGCCTCAGTTTGGTCACGTTCTCT
GATGGGTGCATTTAGATTGTGTTCAGGTTTGGAGGCATTAAAATTCTGCATAGCTCTGAC
CCTGAGCTTGACCTAATTGTAGAATGTCGCCTTGCTGGTGAAAAAGGAAAGACATTTTCT
ATAATGAATCTGTAAATCTGATCATGTACGAATGTCTGACTGTTTTATATTTTATTTTAG
TGACTCACATCATTAAACATTTTAAAGTGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAA

SEQ ID NO:2

MSVMPSNLSLNGTSFFAENHSIMDKPNEQRTLNVFLFCSTFIIAFTVLLGSIYSLVSLLKLQNKSTISMIVT
SLSIDDLISIVPVIIFMLTQWSSDALPQPLCTTSALIYLFQGISSNLKGSLIVSYNFYSINKTETMNCSASK
RRVSMVWAILSIWIVSLLICILPLCGWGKYIPTTWGCFTDHASSYILFLFIVYSLCFCLLTVLSVPLTYQLL

CSDEQQLLHVNYQEISRGYITPGTPAGCSTATPCLSPVDPVDKTLKHFQNACLSSKAVFRKGVAESSGLEPR
CMNNAKSRSFTVGFAQKRFSLILALTKVILWLPMMIQMVVQHITGYQSFSFETLSFLLTLLAATVTPVFVLS
EHWIHLPCGCIINCRRNSYAVSSEELKTKHRGFEFNLSFQHGYGIYRISPESHHHDGDGKSTSCHNLLVCEK
PYEPPRGGGGGAALAELSTTDSARPGPAGLSATDSARPGPAGFSTTDSARPSAAGVRGEAASGLGRTVEGPE
RRLSHEEGHKPELTDWEWCRSKSERTPRQRSGGALAIPLCAFQGTVSLQAPTGKTLSLSTYEVSTEGQKITP
TSKKIEVYRSKSVGHDPNPEECPNTFADTSVKIHLEVLEICDNEEALDTVSIISNISQSSTQVRSPSLRYSR
KENRFVSCDLGETASYSLFIPSNNPDSDINITIPDTVEAHRQNSKKQHMERGGYQEEIQMLNKAYRKREEDG
NSN

SEQ ID NO:3

ATGTCTTTCTTTCTCAGCAACTTAACAAATGACTCTAACCTGTGGAAAGCAAGTCATAAT
TCTACGGAGACGGACCTTATGAATTCACCAGCAACTCTGACTCTTTCTCTCTTTTGCCTG
ATCTGTATAATGACTCTCGCAGCTCTGGTGGGCAGCATTTTCTCACTGGTTTCGCTGCTG
ACCATGCAATACAGAACTGTTTTGTCCATACTTGTGACATCTTGGTCTGTGGATGATCTC
TTGAGCGTCTTGTCCGTGGCCATCTTCATGGTTTTACAGTGGCCAAAAGAGGCACAGGGC
TACTTCCAGTCTCTGTGCACCACCTCTGCCCTACTGTATATGTGCCAGGGCCTCTCCAGC
AACTTGAAGGCGACTCTCATAGTCTGCTACAACTTTTATACGATGAACAGGACGGTGGAG
AGTCAGTCAAGCTCCTGGCGATTAGGACAGGTACTTGGTGTGACTTTAACCGTTTGGGCA
GTCAGCCTGCTGCTGGCTTCGCTCCCGCTCTGTGGCTGGGGTGTCTTCGTGCGCACGCCT
TGGGGCTGCCTGACCGACTGCTCCAGCCCATACGTGCTGCTCCTCTTCGCGGTATACGCT
TCCGCTTTCGGGCTCCTGGCGGTTCTCTCTGTCCCTCTCACTCACCAGCTGCTGTGTTCG
GAGGAACCGCTGAGACTCCATGCCAACTACCAGGAAATTTCTCGTGGCGCCTCTACTCCT
GGGACCCCTGCTGCTGCGGGAAGAGTGCTCTGTCTGCCGCCAGAGGATGTTGAGATCCCT
GCTCTGCGATGCACTGGGGGATGCTCCCCGAGCTCCGACGTGGTGTTTGCCCCGGGTCAG
CCGGCTGCCTCCGGCGCTGGAGCCGGCAGGCGAGAGAATCCCGGGACTCCCCAGGGCACC
AACAGCTTCCCCTTGAGCTTGGCACAGAAGCGCTTTTCTTTGATTCTAGCGCTGACGAAA
GTCATCCTTTGGCTACCCATGATGATCCACATGGTGGTAAAACACGTGGTGGGGTTTCAG
AGCCTTCCGGTCGATATGCTCAGCTTTCTACTAAGCCTGCTGGCCAGCAGTGTAACACCA
GTGTTTGTCTTATCCAAACGTTGGGCCCACTTGCCTTGTGGCTGCATCATCAACTGCCAG
CCAGACACCTACTCCGTTGCCTTCGATGGGAAAAAGTCGAAGAGGAAAGGCTTTGAATTC
AATCTATCATTCCAACAAAGTTATGGACTCTATAAAATACCACATGCAGATTGCTATGAT
GATGATGAAAATTCCATATCCTATCACAACCCAAAGAATTATGAGTGTGAAGCTACAAAG
GAGCCTTGGGGAGACAACCGCAGTGTCTTCAATACCATCACAGTAGAAATCAGCACCACG
CCACCGCTGGACAGCGCCACACTGACGGGTGTCCACAAGTGTACCAATACTGACATCCCA
GAATCCAAACAGGCCATGAAAGAGGAAAAGGGTGCTTTTTCTGTTAAAACAGAAAGCGAT
ATTAACTATGGAGAAACCACCTCTTTTGAAGGTCCAGAGAGACGACTGTCCCATGAGGAG
AATCAGAAACCAGATCTTTCAGACTGGGAATGGTGCAGGAGTAAGTCTGAACGAACGCCT
CGCCAGCGTTCCGGTGGTGGGCTGGCTATTCCCATATGCGCGTTCCAGGGGACTGTGTCT
CTCCAAGCACCTACTGGGAAAACGCTCTCTCTTTCTACTTATGAGGTCAGCGCAGAAGGG
CAAAAGATAACCCCAGCCTCTAAGAAAATAGAAGTCTATCGATCTAAAAGTGTTGGCCAT
GAACCGAACTCAGAGGAGTCTCCCTCCACGTTTGCGGACACCAGTGTAAAAATACACTTG
GAAGTTCTTGAAATTTGTGACAACGATGAGGCTCTAGACACTGTGTCAATTATCAGCAAC
ATCAGCCAATCCTCCACAAAAGTCAGGTCTCCTTCTCTGCGCTACTCACGGAAGGAAAAC
AGATTCGTCTCCTGCGACTTAGGAGAAACTGCCTCATACTCGCTCTTTTTACCCACAAGT
GATCCTGATGGTGATATTAACATCTCTATCCCTGACACGGTGGAGGCCCACAGACAGAAC
AGCAGGCGGCAGCATGAAGAGAGGGATGGCTACCAGGAAGAGATCCAGTTGTTAAACAAA
GCCTACAGGAAAAGAGAGGCAGAAAGCAAGGGTGATTAG

SEQ ID NO:4

MSFFLSNLTNDSNLWKASHNSTETDLMNSPATLTLSLFCLICIMTLAALVGSIFSLVSLL
TMQYRTVLSILVTSWSVDDLLSVLSVAIFMVLQWPKEAQGYFQSLCTTSALLYMCQGLSS
NLKATLIVCYNFYTMNRTVESQSSSWRLGQVLGVTLTVWAVSLLLASLPLCGWGVFVRTP
WGCLTDCSSPYVLLLFAVYASAFGLLAVLSVPLTHQLLCSEEPLRLHANYQEISRGASTP
GTPAAAGRVLCLPPEDVEIPALRCTGGCSPSSDVVFAPGQPAASGAGAGRRENPGTPQGT
NSFPLSLAQKRFSLILALTKVILWLPMMIHMVVKHVVGFQSLPVDMLSFLLSLLASSVTP
VFVLSKRWAHLPCGCIINCQPDTYSVAFDGKKSKRKGFEFNLSFQQSYGLYKIPHADCYD
DDENSISYHNPKNYECEATKEPWGDNRSVFNTITVEISTTPPLDSATLTGVHKCTNTDIP
ESKQAMKEEKGAFSVKTESDINYGETTSFEGPERRLSHEENQKPDLSDWEWCRSKSERTP
RQRSGGGLAIPICAFQGTVSLQAPTGKTLSLSTYEVSAEGQKITPASKKIEVYRSKSVGH
EPNSEESPSTFADTSVKIHLEVLEICDNDEALDTVSIISNISQSSTKVRSPSLRYSRKEN
RFVSCDLGETASYSLFLPTSDPDGDINISIPDTVEAHRQNSRRQHEERDGYQEEIQLLNK
AYRKREAESKGD

SEQ ID NO:5

ATGTCGTTCTTCCTCAGCAACTTAACAAATGACTCTAGACTGTGGAAAGTAAGTCATAAT
TCTACGGACCTTATGAATTCGCCAGAAACTCTGACTCTTTCTCTCTTTTGCCTGATCTGT
TTAATGACTCTCGTAGCTCTGGTAGGCAGCATTTTCTCACTGGTTTCTCTGCTCACCATG
CAGTACAGAACTGTTGTGTCTATGCTTGTGACCTCTTGGTCTGTGGATGATCTCTTGAGC
GTCTTGTCAGTGGCCATCTTCATGGTTTTGCAGTGGCCAAGAGAGGCACCGGGCTACTTC
CAGTCTCTGTGTACCACCTCTGCCTTACTGTATATGTGCCAGGGCCTCTCCAGCAACTTG
AAGGCGACTCTTATAGTCTTCTACAACTTTTATACGATGCACAGGACAGTGGTGAGTCAG
TCAAGCTCCTGGCGATCAGGACAGGTACTCGGTGTGGCTTTAACCGTGTGGGCAGTCAGC
CTGCTGCTGGCCTCGCTCCCGCTGTGCGGCTGGGGTGTCTTCGTGCGCACGCCTTGGGGC
TGCCTGACCGACTGCTCCAGCCCATACGTGCTGCTCCTCTTCGCGGTGTACGCTTCCGCT
TTCGGGCTCCTGGCGGTTCTTTCTGTCCCTCTCACTCACCAGCTGCTGTGTTCAGAGGAG
CCGCCAAGACTCCATGCCAACTACCAGGAAATTTCTCGTGGCGCCTCTACTCCTGGGACC
CCTGCTGCTGGGGGAAGAGTGCTCTGTCTGTTGCCAGAGGATGTTGAAATCCCTGCTCTG
CCAGGCACTGGGAGCTCCCTGAGCTCCGACATGGTGTTTGCACCGGGTCAGCCTGCTGCC
TCTAGCGCTGGAGCCGGCAAGCGAGAGAATCTCTGGACTCCCCGTGGCTCGAGCAGCTTC
CCCGTGAGCCTGGCACAGAAGCGCTTTGCTTTGATCCTAGCGCTGACTAAAGTCATCCTC
TGGCTGCCTATGATGATCCACATGGTGGTAAAACACGTGGTGGGGTTTCAGAGCCTTCCG
GTCGATATGCTCAGCTTTCTACTAACCCTGCTGGCCAGCACTGTAACACCAGTGTTTGTC
TTGTCTAAACGATGGGCCCACTTGCCTTGTGGCTGCATCATCAACTGCCAGCCTGACACC
TACTCCGTTGCATTCGATGGAAAAAGTCAAAGAGGAAAGGCTTTGAATTCAATCTATCG
TTCCAACAAAGTTATGGACTCTATAAAATGACACATGCAGATTACTATGACGATGATGAT
GAAAATCCCATATCCTACCACAACCCAAAGAAATATGAGTGTGAAGCTACAAAGGAGCCT
CGGGAAGACAACCACGGTGTGTTCAATACCATCACAGTAGAAATCAGCACCACGCCACCG
CTGGACAGCGCCACACTGACCGGTGTCAACAAGTGCACCAACACGGATATCCCTGAGCCC
AAACAGGCCGTGAGCGAAGAAAAGGGTGCCTTTTCTATTAAAACAGAATGTGCTATTAAC
TATGGAGAAGCCACCTCGTTTGAAGGTCCAGAGAGACGACTGTCCCATGAGGAGACTCAG
AAACCAGATCTTTCAGACTGGGAGTGGTGCAGGAGTAAGTCAGAACGAACACCTCGCCAG
CGTTCTGGTGGTGGGCTGGCTATTCCCATTTGTGCGTTCCAAGGGACTGTGTCTCTCCAA
GCACCAACTGGGAAAACCCTCTCTCTTTCTACTTATGAGGTCAGTGCAGAAGGGCAAAAG
ATAACCCCGCCCTCTAAGAAAATAGAAGTCTATCGATCTAAAAGTGTTGGCCATGAACCG
AACTCAGAAGAGTCTCCCTCCACGTTTGCAGACACCAACGTAAAAATACACTTGGAAGTT

CTTGAAATTTGTGACAACGATGAGGCTTTAGACACTGTGTCAATCATCAGCAACATCAGC
CAGTCCTCTACAAAAGTTCGGTCTCCTTCTCTGCGCTACTCACGGAAGGAAAACAGGTTT
GTCTCATGCGACTTAGGAGAAACTGCCTCCTACTCCCTCTTCTTACCCACAAGTGATCCT
GACGGAGATATCAACATCTCTATCCCAGACACAGTGGAGGCACACAGACAGAACAGCAGG
CGGCAGCATCAAGACAGGGATGGCTACCAGGAAGAGATCCAGTTGTTAAACAAAGCCTAC
AGGAAAAGAGAAGCAGAAAGCAAGGGTAACTAG

SEQ ID NO:6

MSFFLSNLTNDSRLWKVSHNSTDLMNSPETLTLSLFCLICLMTLVALVGSIFSLVSLLTM
QYRTVVSMLVTSWSVDDLLSVLSVAIFMVLQWPREAPGYFQSLCTTSALLYMCQGLSSNL
KATLIVFYNFYTMHRTVVSQSSSWRSGQVLGVALTVWAVSLLLASLPLCGWGVFVRTPWG
CLTDCSSPYVLLLFAVYASAFGLLAVLSVPLTHQLLCSEEPPRLHANYQEISRGASTPGT
PAAGGRVLCLLPEDVEIPALPGTGSSLSSDMVFAPGQPAASSAGAGKRENLWTPRGSSSF
PVSLAQKRFALILALTKVILWLPMMIHMVVKHVVGFQSLPVDMLSFLLTLLASTVTPVFV
LSKRWAHLPCGCIINCQPDTYSVAFDGKKSKRKGFEFNLSFQQSYGLYKMTHADYYDDDD
ENPISYHNPKKYECEATKEPREDNHGVFNTITVEISTTPPLDSATLTGVNKCTNTDIPEP
KQAVSEEKGAFSIKTECAINYGEATSFEGPERRLSHEETQKPDLSDWEWCRSKSERTPRQ
RSGGGLAIPICAFQGTVSLQAPTGKTLSLSTYEVSAEGQKITPPSKKIEVYRSKSVGHEP
NSEESPSTFADTNVKIHLEVLEICDNDEALDTVSIISNISQSSTKVRSPSLRYSRKENRF
VSCDLGETASYSLFLPTSDPDGDINISIPDTVEAHRQNSRRQHQDRDGYQEEIQLLNKAY
RKREAESKGN

SEQ ID NO:7

GTATGGAGGCATGGAAAGAACAGGTGATTTGCATTCTCACCTCACCAACATAAATCAGGCCTGCCACCATGA
TCAGACGGGCTATGCTGCTGTGACAGTGATGGAGAGCTGAAGAGAATCAGGGTCAGAGACTGCAGGCCGCTG
TGATGGAGATGGATCACTCTGTGCAAGATGTTGCTTTTCACGTGCTGCTCTTCCAAGGTTGAAAACAAAGTT
CCTCCATAAGATTGCCTCAAGGCTCCTTAGTA

SEQ ID NO:8

GTGTCTTCGTGCGCACGCCTTGGGGCTGCCTGACCGACTGCTCCAGCCCATACGTGCTGC
TCCTCTTCGCGGTATACGCTTCCGCTTTCGGGCTCCTGGCGGTTCTCTCTGTCCCTCTCA
CTCACCAGCTGCTGTGTTCGGAGGAACCGCTGAGACTCCATGCCAACTACCAGGAAATTT
CTCGTGGCGCCTCTACTCCTGGGACCCCTGCTGCTGCGGGAAGAGTGCTCTGTCTGCCGC
CAGAGGATGTTGAGATCCCTGCTCTGCGATGCACTGGGGGATGCTCCCCGAGCTCCGACG
TGGTGTTTGCCCCGGGTCAGCCGGCTGCCTCCGGCGCTGGAGCCGGCAGGCGAGAGAATC
CCGGGACTCCCCAGGGCACCAACAGCTTCCCCTTGAGCTTGGCACAGAAGCGCTTTTCTT
TGATTCTAGCG

SEQ ID NO:9

VFVRTPWGCLTDCSSPYVLLLFAVYASAFGLLAVLSVPLTHQLLCSEEPLRLHANYQEIS
RGASTPGTPAAAGRVLCLPPEDVEIPALRCTGGCSPSSDVVFAPGQPAASGAGAGRRENP
GTPQGTNSFPLSLAQKRFSLILA

SEQ ID NO:10

ATGTCTTTATTTCTCAGTAACTTATCAACAAATGACTCTAGCCTGTGGAAAGAGAATCAT
AATTCTACGGACCTTTTAAATCCGCCAGGAACCCTGAATATCTATCTTTTTTGCTTGACA
TGTCTCATGACTTTTGCAGCCTTGGTGGGCAGCATTTATTCACTAATTTCCCTGCTGAAA
ATGCAGAACAGAACTGTTGTGTCCATGCTTGTGGCTTCCTGGTCTGTGGATGATCTCATG
AGCGTCCTGTCGGTGACCATCTTCATGTTTTTGCAGTGGCCAAACGAGGTCCCCGGTTAC
TTCCAATTTCTGTGCACCACCTCTGCCTTAATGTATTTATGCCAGGGCCTCTCTAGCAAC
TTGAAGGCGACTCTCCTAGTCTCTTACAACTTTTATACGATGCACAGAGGTGTGGGGAGC
CAGACAGCCTCCAGAAGATCGGGCCAGGTGCTCGGCGTGGTGCTGACCGTGTGGGCAGCC
AGTCTGCTGCTCTCGGCGCTCCCGCTGTGCGGCTGGGGCGCCTCCGTGCGCACGCCCTGG
GGCTGCCTGGTGGACTGCTCCAGCTCCTACGTACTATTCCTCTCTATCGTGTACGCTTTG
GCCTTCGGACTCCTCGTGGGCCTCTCAGTCCCACTCACTCACCGATTGCTGTGTTCGGAG
GAGCCGCCGAGACTCCACTCCAACTACCAGGAAATTTCCCGTGGAGCTTCAATTCCTGGG
ACCCCTCCTACTGCGGGGAGAGTGGTTTCCCTGTCCCCAGAGGATGCTCCAGGCCCGAGT
CTGCGGCGCTCTGGGGGATGCTCTCCGAGCTCCGACACCGTGTTCGGACCGGGTGCGCCC
GCTGCCGCTGGGGCTGAAGCCTGCAGGCGTGAGAACCGGGGGACTCTCTATGGCACCAGG
AGCTTCACCGTGAGCGTAGCGCAGAAGCGCTTCGCTTTGATCCTAGCGCTTACAAAAGTC
GTCCTTTGGCTGCCCATGATGATGCACATGGTGGTCCAGAACGTCGTGGGGTTTCAGAGC
CTTCCCTTGGAGACATTCAGCTTCCTACTTACCCTGCTGGCCACCACTGTAACCCCAGTG
TTTGTCTTGTCCAAACGCTGGACCCACTTGCCCTGTGGCTGCATCATCAACTGCAGGCAG
AATGCATATGCAGTGGCGTCCGATGGAAAAAAATCAAGAGAAAAGGCTTTGAATTCAAT
CTATCATTCCAAAAAAGTTATGGGATTTATAAAATAGCACATGAAGATTACTATGATGAT
GATGAAAATTCCATATTCTATCACAACCTGATGAACTCTGAGTGTGAAACTACAAAAGAC
CCTCAGAGAGACAACCGTAACATCTTCAATGCTATAAAAGTAGAAATCAGCACCACGCCC
TCTCTGGACAGCTCCACACAAAGAGGCATCAACAAATGCACAAATACTGATATTACAGAA
GCTAAACAGGATTCCAACAACAAAAAGGATGCGTTTTCTGACAAAACAGGAGGTGATATT
AACTATGAAGAAACTACCTTTTCTGAAGGGCCAGAAAGAAGACTGTCTCATGAAGAGAGT
CAGAAACCAGATCTTTCAGACTGGGAGTGGCGTAGGAGTAAATCAGAAAGAACCCCTCGT
CAGCGTTCCGGTTATGCCCTTGCCATTCCCTTGTGTGCATTCCAGGGGACTGTGTCTCTC
CATGCACCTACAGGGAAAACCCTATCTCTTTCTACCTATGAGGTAAGCGCAGAAGGGCAA
AAAATAACTCCAGCCTCTAAGAAAATAGAAGTCTATCGATCCAAAAGTGTTGGCCATGAA
CCAAACTCAGAAGATTCTTCATCC

SEQ ID NO:11

MSLFLSNLSTNDSSLWKENHNSTDLLNPPGTLNIYLFCLTCLMTFAALVGSIYSLISLLK
MQNRTVVSMLVASWSVDDLMSVLSVTIFMFLQWPNEVPGYFQFLCTTSALMYLCQGLSSN
LKATLLVSYNFYTMHRGVGSQTASRRSGQVLGVVLTVWAASLLLSALPLCGWGASVRTPW
GCLVDCSSSYVLFLSIVYALAFGLLVGLSVPLTHRLLCSEEPPRLHSNYQEISRGASIPG
TPPTAGRVVSLSPEDAPGPSLRRSGGCSPSSDTVFGPGAPAAAGAEACRRENRGTLYGTR
SFTVSVAQKRFALILALTKVVLWLPMMMHMVVQNVVGFQSLPLETFSFLLTLLATTVTPV
FVLSKRWTHLPCGCIINCRQNAYAVASDGKKIKRKGFEFNLSFQKSYGIYKIAHEDYYDD
DENSIFYHNLMNSECETTKDPQRDNRNIFNAIKVEISTTPSLDSSTQRGINKCTNTDITE

AKQDSNNKKDAFSDKTGGDINYEETTFSEGPERRLSHEESQKPDLSDWEWRRSKSERTPR
QRSGYALAIPLCAFQGTVSLHAPTGKTLSLSTYEVSAEGQKITPASKKIEVYRSKSVGHE
PNSEDSSS

Claims

1. An isolated polypeptide comprising a polypeptide sequence which has at least 95% identity to the polypeptide sequence of SEQ ID NO:2, 4, 6 or 11.

2. The polypeptide as claimed in claim 1 comprising the polypeptide sequence of SEQ ID NO:2, 4, 6 or 11.

3. The isolated polypeptide of SEQ ID NO:2, 4, 6 or 11.

4. An isolated polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence which has at least 95% identity to a polypeptide of SEQ ID NO:2, 4, 6 or 11; or a polynucleotide sequence complementary to said isolated polynucleotide.

5. An isolated polynucleotide which comprises a polynucleotide sequence which has at least 95% identity to that of SEQ ID NO:1, 3, 5 or 10; or a polynucleotide sequence complementary to said isolated polynucleotide.

6. An isolated polynucleotide selected from:
(a) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2, 4, 6 or 11;
(b) a polynucleotide encoding the polypeptide of SEQ ID NO:2, 4, 6 or 11;
(c) the polynucleotide of SEQ ID NO:1, 3, 5 or 10; and
(d) a polynucleotide obtainable by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1, 3, 5, 10 or a fragment thereof;
or a nucleotide sequence complementary to said isolated polynucleotide over the entire length thereof.

7. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.

8. A host cell comprising the expression system of claim 7 or a membrane thereof expressing the polypeptide of claim 1.

9. A process for producing a polypeptide of claim 1 comprising the step of culturing a host cell of claim 8 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.

11. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:
(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;
(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presense of a labeled competitior;
(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;
(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring the activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or
(e) detecting the effect of a candidate compound on the production ofmRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

12. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the polypeptide of claim 1 in a subjectcomprising:
(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or
(b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

SEQUENCE LISTING

<110> SmithKline Beecham plc
Max-Planck Institute of Neurobiology

<120> Novel Compounds

<130> P32290

<160> 11

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 4683
<212> DNA
<213> Gallus gallus

<400> 1

```
ctaagaggca gctctgtctg ggaagatact cactccccat ttctctgcta tcctgcttgc      60
ttctgactcc tccctcctgc agctcagatc agcacagggt gaatccaaca gcctcccatt     120
tctaccgtgt ttctccctga gcttgagcat acctggtact ctgccttcgt ttcagctgaa     180
ctcgtgagta aggcacattg cctgatatct gtcaaaagtg aatggcttgt tattacctga     240
agagactgaa ttaactatct tctggtactg aaagggtaat gaagaactgt attttttcca     300
gcacctttg tacttctttg ttgcctgatt tttcaagagc gtgagttatt ctggttatca      360
agggagctta aacctgagta tttcttcacc ggatgctaca aaagaagaaa ggaaaacaaa     420
cttaagatgt ggagctagca gttatggttt cagcttgtga aaacttgcag agtcagtgag     480
aaggcaagga acaacatttg ccaaggacca tgtcggtaat gcctagtaat ttgtccctca     540
atgggacaag cttcttcgct gaaaatcata gcattatgga taaacctaat gagcagagaa     600
ctttgaatgt ctttctattt tgctcgactt ttatcattgc attcacagtg cttctgggca     660
gcatttattc actagtttcc ctgctgaaac tgcagaacaa atccacaatt tcaatgattg     720
tgacctcttt gtcaatagat gatttgatca gcattgtgcc agtgattatt ttcatgctca     780
cccagtggtc aagtgatgcc cttcctcagc ctctgtgcac cacctcagca cttatatatt     840
tattccaggg catttctagc aatctgaaag ggtctcttat agtttcctac aacttctata     900
gcatcaacaa aactgagaca atgaactgca gtgcttccaa acgacgagtg agcatggtat     960
gggcaattct tagcatttgg atcgtcagtt tgctgatatg cattttgcct ctctgtggct    1020
ggggcaaata catccccacc acctggggtt gcttcacaga ccatgcaagt tcctacatct    1080
tgttcttatt tattgtctac tcattgtgct tttgcctcct cacagtgctg tctgtccctc    1140
taacttacca gctgttgtgc tcagatgagc aacagctatt acacgtcaac taccaggaaa    1200
tctctcgagg ctacatcacc cctgggacac ctgcaggctg cagtactgcc accccatgtc    1260
tgtcaccagt ggaccctgtg gataaaaccc tgaagcattt ccaaaatgca tgtctaagct    1320
```

caaaggcagt ttttaggaaa ggtgtggctg agagcagtgg actcgagccc cgatgcatga 1380
acaacgcaaa gagcaggagc ttcacagtgg gctttgccca gaaacggttc tcactgattc 1440
ttgcattgac aaaagtcatt ctctggctcc caatgatgat acaaatggtt gtccagcaca 1500
tcactggcta tcaaagcttt tcatttgaga cacttagctt cctgctgaca ttgctagctg 1560
ccaccgtcac cccagtattt gtcctgtcag agcactggat ccacctgccg tgtggctgca 1620
tcataaactg caggaggaat tcatatgctg tgtcttcaga agaactcaaa accaagcata 1680
gaggtttcga attcaacctt tccttccagc atggttatgg aatctaccga atatccccg 1740
agagccacca ccacgatggc gatggtaaat ccacatcgtg ccacaacctg ctggtttgcg 1800
agaaaccgta cgagccgccg aggggaggtg gcggcggtgc ggcgctggcg gaactcagca 1860
ccacggatag cgcccggcca ggcccggcgg ggctcagcgc cacggacagc gccaggcccg 1920
gtcccgcggg gttcagcacc acagacagcg ccaggcccag cgccgcgggg gtgcgcgggg 1980
aggcggcctc gggcctcggc aggaccgtgg agggacccga gaggaggctg tcgcacgaag 2040
agggccataa gccagagctc acagactggg agtggtgcag gagtaaatca gagaggaccc 2100
ctcgacagcg gtcaggtggc gccttagcta ttcctctgtg tgcatttcaa ggaaccgtgt 2160
cacttcaggc acccacagga aaaactctct ccttgtctac atacgaggta agcactgaag 2220
gacagaaaat aacacccaca tcaaagaaaa ttgaagttta tcgatctaaa agtgttggtc 2280
atgaccccaa cccagaggag tgtcctaata cgtttgctga cacgagtgtt aaaattcatt 2340
tagaggttct ggaaatttgt gacaatgaag aggctctgga cactgtttca atcatcagta 2400
atatcagcca gtcctccaca caggtaagat ctccatcctt acgctactca cggaaggaga 2460
acagatttgt ctcatgtgat ttgggggaaa ctgcttccta ctcactcttc ataccatcca 2520
acaatcccga cagcgatatt aacataacaa ttccagacac tgtggaagct caccggcaga 2580
atagtaaaaa gcaacacatg gaaagaggtg gttatcagga agaaatacag atgttgaata 2640
aagcgtacag aaaacgggaa gaagatggca acagcaatta aaagacattt aatctaaaca 2700
tttaaactga ttcagcagac attgctcctc ttgactcaag ttaccctaac aaatacaatg 2760
ccaaccggtt tgtttttttt tttttcctag catataaatg atttttattt gggtacactg 2820
ttaaatagtt aacatttaca aatgggaaaa caaacgcttc agtgctttgc agctttgtgc 2880
aatgaaatcc aagttattta ccaatctctt tatttgagga agacttgcat ctgaaaggaa 2940
tcctaaatga actcttaaaa gccacccttg cttgcttgtt ggcactgggt ggtctgtggg 3000
aagccagaaa atgagagagc agctctacac ccctgaagat gtaacaggtt tcaatgactg 3060
tggcagttct cctactgatg tgccagactt gaaacattta ctgaaaatat gtagcacaag 3120
cctaagacta agccatgcat ttggttaggg attgtaagtt acctacgtgg tttaattttc 3180
tactgctcca tacgtaattg atacgtaatg tctctgtaaa tggccattgt atagaccttt 3240
gaatgtgggt gtctgaaatt taagcttgca cattttatat gtccagcact gtgaaaatgt 3300
gttctcatgg ctggacattt ctttcatcga gtgtgtcaaa tgttggttga tcttatgctg 3360
gttactctaa agtgtgccat ttctgcctc cttgatctct taggagaggg gacagggctg 3420
ctcaaggagg ccaaacttag agacccagct tagctgttga cctcagcttg aggctgagtg 3480
cccagatcag cttccctgcc ctggagcacc tctggaggag cctgttctta tgctgcctta 3540
caggaagcgt agagccaagt ctgaaatcaa actataaccc accaatagac tgcttcttgt 3600
caaattcaaa cagactgagg gtaaaacagc aaatgaaaca cttaagtcta atgctttttg 3660
tctgagagct cattttctga gatgttgtca tacatttcag ccagatagaa gtagacttca 3720
gcaggacttt aagaggtttg actaaaagtt acaccacttc tttaattaat agatttaatt 3780
ttcagtgagt gacaacataa taattcaggt gtctgcagtc accttgttat ccattcagtg 3840
ggaacatgca caggaaagta agcacagccc tatttaggat tcctttatgg attgtgaaca 3900

ggtagagcaa aagggccaac tagctcagct tgtgttcagt atggaggcat ggaaagaaca 3960
ggtgatttgc attctcacct caccaacata aatcaggcct gccaccatga tcagacgggc 4020
tatgctgctg tgacagtgat ggagagctga agagaatcag ggtcagagac tgcaggccgc 4080
tgtgatggag atggatcact ctgtgcaaga tgttgctttt cacgtgctgc tcttccaagg 4140
ttgaaaacaa agttcctcca taagattgcc tcaaggctcc ttagtacctg gtgatgctgc 4200
gttacacttc agatttatca agaaactata gtagacagag caatcaatgg taagtgtatt 4260
agtaggttat atccatgtgc atgcagcagg gagacacgtt ctgctgtctc tgtattactt 4320
tgctttgtgt ttgtaggatt agctcagaaa gttgctttat agcaacagat gtgctcaatc 4380
cagaaagacg ttgtgaggga ggatgcagtg catcactgcc tcagtttggt cacgttctct 4440
gatgggtgca tttagattgt gttcaggttt ggaggcatta aaattctgca tagctctgac 4500
cctgagcttg acctaattgt agaatgtcgc cttgctggtg aaaaaggaaa gacattttct 4560
ataatgaatc tgtaaatctg atcatgtacg aatgtctgac tgttttatat tttattttag 4620
tgactcacat cattaaacat tttaaagtgt caaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4680
aaa 4683

<210> 2
<211> 723
<212> PRT
<213> Gallus gallus

<400> 2

Met Ser Val Met Pro Ser Asn Leu Ser Leu Asn Gly Thr Ser Phe Phe
1 5 10 15
Ala Glu Asn His Ser Ile Met Asp Lys Pro Asn Glu Gln Arg Thr Leu
20 25 30
Asn Val Phe Leu Phe Cys Ser Thr Phe Ile Ile Ala Phe Thr Val Leu
35 40 45
Leu Gly Ser Ile Tyr Ser Leu Val Ser Leu Leu Lys Leu Gln Asn Lys
50 55 60
Ser Thr Ile Ser Met Ile Val Thr Ser Leu Ser Ile Asp Asp Leu Ile
65 70 75 80
Ser Ile Val Pro Val Ile Ile Phe Met Leu Thr Gln Trp Ser Ser Asp
85 90 95
Ala Leu Pro Gln Pro Leu Cys Thr Thr Ser Ala Leu Ile Tyr Leu Phe
100 105 110
Gln Gly Ile Ser Ser Asn Leu Lys Gly Ser Leu Ile Val Ser Tyr Asn
115 120 125
Phe Tyr Ser Ile Asn Lys Thr Glu Thr Met Asn Cys Ser Ala Ser Lys
130 135 140
Arg Arg Val Ser Met Val Trp Ala Ile Leu Ser Ile Trp Ile Val Ser
145 150 155 160
Leu Leu Ile Cys Ile Leu Pro Leu Cys Gly Trp Gly Lys Tyr Ile Pro
165 170 175

Thr Thr Trp Gly Cys Phe Thr Asp His Ala Ser Ser Tyr Ile Leu Phe
180 185 190
Leu Phe Ile Val Tyr Ser Leu Cys Phe Cys Leu Leu Thr Val Leu Ser
195 200 205
Val Pro Leu Thr Tyr Gln Leu Leu Cys Ser Asp Glu Gln Gln Leu Leu
210 215 220
His Val Asn Tyr Gln Glu Ile Ser Arg Gly Tyr Ile Thr Pro Gly Thr
225 230 235 240
Pro Ala Gly Cys Ser Thr Ala Thr Pro Cys Leu Ser Pro Val Asp Pro
245 250 255
Val Asp Lys Thr Leu Lys His Phe Gln Asn Ala Cys Leu Ser Ser Lys
260 265 270
Ala Val Phe Arg Lys Gly Val Ala Glu Ser Ser Gly Leu Glu Pro Arg
275 280 285
Cys Met Asn Asn Ala Lys Ser Arg Ser Phe Thr Val Gly Phe Ala Gln
290 295 300
Lys Arg Phe Ser Leu Ile Leu Ala Leu Thr Lys Val Ile Leu Trp Leu
305 310 315 320
Pro Met Met Ile Gln Met Val Val Gln His Ile Thr Gly Tyr Gln Ser
325 330 335
Phe Ser Phe Glu Thr Leu Ser Phe Leu Leu Thr Leu Leu Ala Ala Thr
340 345 350
Val Thr Pro Val Phe Val Leu Ser Glu His Trp Ile His Leu Pro Cys
355 360 365
Gly Cys Ile Ile Asn Cys Arg Arg Asn Ser Tyr Ala Val Ser Ser Glu
370 375 380
Glu Leu Lys Thr Lys His Arg Gly Phe Glu Phe Asn Leu Ser Phe Gln
385 390 395 400
His Gly Tyr Gly Ile Tyr Arg Ile Ser Pro Glu Ser His His His Asp
405 410 415
Gly Asp Gly Lys Ser Thr Ser Cys His Asn Leu Leu Val Cys Glu Lys
420 425 430
Pro Tyr Glu Pro Pro Arg Gly Gly Gly Gly Gly Ala Ala Leu Ala Glu
435 440 445
Leu Ser Thr Thr Asp Ser Ala Arg Pro Gly Pro Ala Gly Leu Ser Ala
450 455 460
Thr Asp Ser Ala Arg Pro Gly Pro Ala Gly Phe Ser Thr Thr Asp Ser
465 470 475 480
Ala Arg Pro Ser Ala Ala Gly Val Arg Gly Glu Ala Ala Ser Gly Leu
485 490 495
Gly Arg Thr Val Glu Gly Pro Glu Arg Arg Leu Ser His Glu Glu Gly
500 505 510
His Lys Pro Glu Leu Thr Asp Trp Glu Trp Cys Arg Ser Lys Ser Glu

515 520 525
Arg Thr Pro Arg Gln Arg Ser Gly Gly Ala Leu Ala Ile Pro Leu Cys
530 535 540
Ala Phe Gln Gly Thr Val Ser Leu Gln Ala Pro Thr Gly Lys Thr Leu
545 550 555 560
Ser Leu Ser Thr Tyr Glu Val Ser Thr Glu Gly Gln Lys Ile Thr Pro
565 570 575
Thr Ser Lys Lys Ile Glu Val Tyr Arg Ser Lys Ser Val Gly His Asp
580 585 590
Pro Asn Pro Glu Glu Cys Pro Asn Thr Phe Ala Asp Thr Ser Val Lys
595 600 605
Ile His Leu Glu Val Leu Glu Ile Cys Asp Asn Glu Glu Ala Leu Asp
610 615 620
Thr Val Ser Ile Ile Ser Asn Ile Ser Gln Ser Ser Thr Gln Val Arg
625 630 635 640
Ser Pro Ser Leu Arg Tyr Ser Arg Lys Glu Asn Arg Phe Val Ser Cys
645 650 655
Asp Leu Gly Glu Thr Ala Ser Tyr Ser Leu Phe Ile Pro Ser Asn Asn
660 665 670
Pro Asp Ser Asp Ile Asn Ile Thr Ile Pro Asp Thr Val Glu Ala His
675 680 685
Arg Gln Asn Ser Lys Lys Gln His Met Glu Arg Gly Gly Tyr Gln Glu
690 695 700
Glu Ile Gln Met Leu Asn Lys Ala Tyr Arg Lys Arg Glu Glu Asp Gly
705 710 715 720
Asn Ser Asn

<210> 3
<211> 2199
<212> DNA
<213> Mus musculus

<400> 3
atgtctttct ttctcagcaa cttaacaaat gactctaacc tgtggaaagc aagtcataat 60
tctacggaga cggaccttat gaattcacca gcaactctga ctctttctct cttttgcctg 120
atctgtataa tgactctcgc agctctggtg ggcagcattt tctcactggt ttcgctgctg 180
accatgcaat acagaactgt tttgtccata cttgtgacat cttggtctgt ggatgatctc 240
ttgagcgtct tgtccgtggc catcttcatg gttttacagt ggccaaaaga ggcacagggc 300
tacttccagt ctctgtgcac cacctctgcc ctactgtata tgtgccaggg cctctccagc 360
aacttgaagg cgactctcat agtctgctac aacttttata cgatgaacag gacggtggag 420
agtcagtcaa gctcctggcg attaggacag gtacttggtg tgactttaac cgtttgggca 480
gtcagcctgc tgctggcttc gctcccgctc tgtggctggg gtgtcttcgt gcgcacgcct 540

tggggctgcc tgaccgactg ctccagccca tacgtgctgc tcctcttcgc ggtatacgct 600
tccgctttcg ggctcctggc ggttctctct gtccctctca ctcaccagct gctgtgttcg 660
gaggaaccgc tgagactcca tgccaactac caggaaattt ctcgtggcgc ctctactcct 720
gggacccctg ctgctgcggg aagagtgctc tgtctgccgc cagaggatgt tgagatccct 780
gctctgcgat gcactggggg atgctccccg agctccgacg tggtgtttgc cccgggtcag 840
ccggctgcct ccggcgctgg agccggcagg cgagagaatc ccgggactcc ccagggcacc 900
aacagcttcc ccttgagctt ggcacagaag cgcttttctt tgattctagc gctgacgaaa 960
gtcatccttt ggctacccat gatgatccac atggtggtaa aacacgtggt ggggtttcag 1020
agccttccgg tcgatatgct cagctttcta ctaagcctgc tggccagcag tgtaacacca 1080
gtgtttgtct tatccaaacg ttgggcccac ttgccttgtg gctgcatcat caactgccag 1140
ccagacacct actccgttgc cttcgatggg aaaaagtcga agaggaaagg ctttgaattc 1200
aatctatcat tccaacaaag ttatggactc tataaaatac cacatgcaga ttgctatgat 1260
gatgatgaaa attccatatc ctatcacaac ccaaagaatt atgagtgtga agctacaaag 1320
gagccttggg gagacaaccg cagtgtcttc aataccatca cagtagaaat cagcaccacg 1380
ccaccgctgg acagcgccac actgacgggt gtccacaagt gtaccaatac tgacatccca 1440
gaatccaaac aggccatgaa agaggaaaag ggtgcttttt ctgttaaaac agaaagcgat 1500
attaactatg gagaaaccac ctcttttgaa ggtccagaga gacgactgtc ccatgaggag 1560
aatcagaaac cagatctttc agactgggaa tggtgcagga gtaagtctga acgaacgcct 1620
cgccagcgtt ccggtggtgg gctggctatt cccatatgcg cgttccaggg gactgtgtct 1680
ctccaagcac ctactgggaa aacgctctct ctttctactt atgaggtcag cgcagaaggg 1740
caaaagataa ccccagcctc taagaaaata gaagtctatc gatctaaaag tgttggccat 1800
gaaccgaact cagaggagtc tccctccacg tttgcggaca ccagtgtaaa aatacacttg 1860
gaagttcttg aaatttgtga caacgatgag gctctagaca ctgtgtcaat tatcagcaac 1920
atcagccaat cctccacaaa agtcaggtct ccttctctgc gctactcacg gaaggaaaac 1980
agattcgtct cctgcgactt aggagaaact gcctcatact cgctcttttt acccacaagt 2040
gatcctgatg gtgatattaa catctctatc cctgacacgg tggaggccca cagacagaac 2100
agcaggcggc agcatgaaga gagggatggc taccaggaag agatccagtt gttaaacaaa 2160
gcctacagga aaagagaggc agaaagcaag ggtgattag 2199

<210> 4
<211> 732
<212> PRT
<213> Mus musculus

<400> 4

Met Ser Phe Phe Leu Ser Asn Leu Thr Asn Asp Ser Asn Leu Trp Lys
1 5 10 15

Ala Ser His Asn Ser Thr Glu Thr Asp Leu Met Asn Ser Pro Ala Thr
20 25 30

Leu Thr Leu Ser Leu Phe Cys Leu Ile Cys Ile Met Thr Leu Ala Ala
35 40 45

Leu Val Gly Ser Ile Phe Ser Leu Val Ser Leu Leu Thr Met Gln Tyr
50 55 60

```
Arg Thr Val Leu Ser Ile Leu Val Thr Ser Trp Ser Val Asp Asp Leu
65                  70                  75                  80
Leu Ser Val Leu Ser Val Ala Ile Phe Met Val Leu Gln Trp Pro Lys
                85                  90                  95
Glu Ala Gln Gly Tyr Phe Gln Ser Leu Cys Thr Thr Ser Ala Leu Leu
            100                 105                 110
Tyr Met Cys Gln Gly Leu Ser Ser Asn Leu Lys Ala Thr Leu Ile Val
        115                 120                 125
Cys Tyr Asn Phe Tyr Thr Met Asn Arg Thr Val Glu Ser Gln Ser Ser
    130                 135                 140
Ser Trp Arg Leu Gly Gln Val Leu Gly Val Thr Leu Thr Val Trp Ala
145                 150                 155                 160
Val Ser Leu Leu Leu Ala Ser Leu Pro Leu Cys Gly Trp Gly Val Phe
                165                 170                 175
Val Arg Thr Pro Trp Gly Cys Leu Thr Asp Cys Ser Ser Pro Tyr Val
            180                 185                 190
Leu Leu Leu Phe Ala Val Tyr Ala Ser Ala Phe Gly Leu Leu Ala Val
        195                 200                 205
Leu Ser Val Pro Leu Thr His Gln Leu Leu Cys Ser Glu Glu Pro Leu
    210                 215                 220
Arg Leu His Ala Asn Tyr Gln Glu Ile Ser Arg Gly Ala Ser Thr Pro
225                 230                 235                 240
Gly Thr Pro Ala Ala Ala Gly Arg Val Leu Cys Leu Pro Pro Glu Asp
                245                 250                 255
Val Glu Ile Pro Ala Leu Arg Cys Thr Gly Gly Cys Ser Pro Ser Ser
            260                 265                 270
Asp Val Val Phe Ala Pro Gly Gln Pro Ala Ala Ser Gly Ala Gly Ala
        275                 280                 285
Gly Arg Arg Glu Asn Pro Gly Thr Pro Gln Gly Thr Asn Ser Phe Pro
    290                 295                 300
Leu Ser Leu Ala Gln Lys Arg Phe Ser Leu Ile Leu Ala Leu Thr Lys
305                 310                 315                 320
Val Ile Leu Trp Leu Pro Met Met Ile His Met Val Val Lys His Val
                325                 330                 335
Val Gly Phe Gln Ser Leu Pro Val Asp Met Leu Ser Phe Leu Leu Ser
            340                 345                 350
Leu Leu Ala Ser Ser Val Thr Pro Val Phe Val Leu Ser Lys Arg Trp
        355                 360                 365
Ala His Leu Pro Cys Gly Cys Ile Ile Asn Cys Gln Pro Asp Thr Tyr
    370                 375                 380
Ser Val Ala Phe Asp Gly Lys Lys Ser Lys Arg Lys Gly Phe Glu Phe
385                 390                 395                 400
Asn Leu Ser Phe Gln Gln Ser Tyr Gly Leu Tyr Lys Ile Pro His Ala
```

405 410 415

Asp Cys Tyr Asp Asp Asp Glu Asn Ser Ile Ser Tyr His Asn Pro Lys

420 425 430

Asn Tyr Glu Cys Glu Ala Thr Lys Glu Pro Trp Gly Asp Asn Arg Ser

435 440 445

Val Phe Asn Thr Ile Thr Val Glu Ile Ser Thr Thr Pro Pro Leu Asp

450 455 460

Ser Ala Thr Leu Thr Gly Val His Lys Cys Thr Asn Thr Asp Ile Pro

465 470 475 480

Glu Ser Lys Gln Ala Met Lys Glu Glu Lys Gly Ala Phe Ser Val Lys

485 490 495

Thr Glu Ser Asp Ile Asn Tyr Gly Glu Thr Thr Ser Phe Glu Gly Pro

500 505 510

Glu Arg Arg Leu Ser His Glu Glu Asn Gln Lys Pro Asp Leu Ser Asp

515 520 525

Trp Glu Trp Cys Arg Ser Lys Ser Glu Arg Thr Pro Arg Gln Arg Ser

530 535 540

Gly Gly Gly Leu Ala Ile Pro Ile Cys Ala Phe Gln Gly Thr Val Ser

545 550 555 560

Leu Gln Ala Pro Thr Gly Lys Thr Leu Ser Leu Ser Thr Tyr Glu Val

565 570 575

Ser Ala Glu Gly Gln Lys Ile Thr Pro Ala Ser Lys Lys Ile Glu Val

580 585 590

Tyr Arg Ser Lys Ser Val Gly His Glu Pro Asn Ser Glu Glu Ser Pro

595 600 605

Ser Thr Phe Ala Asp Thr Ser Val Lys Ile His Leu Glu Val Leu Glu

610 615 620

Ile Cys Asp Asn Asp Glu Ala Leu Asp Thr Val Ser Ile Ile Ser Asn

625 630 635 640

Ile Ser Gln Ser Ser Thr Lys Val Arg Ser Pro Ser Leu Arg Tyr Ser

645 650 655

Arg Lys Glu Asn Arg Phe Val Ser Cys Asp Leu Gly Glu Thr Ala Ser

660 665 670

Tyr Ser Leu Phe Leu Pro Thr Ser Asp Pro Asp Gly Asp Ile Asn Ile

675 680 685

Ser Ile Pro Asp Thr Val Glu Ala His Arg Gln Asn Ser Arg Arg Gln

690 695 700

His Glu Glu Arg Asp Gly Tyr Gln Glu Glu Ile Gln Leu Leu Asn Lys

705 710 715 720

Ala Tyr Arg Lys Arg Glu Ala Glu Ser Lys Gly Asp

725 730

<210> 5

<211> 2193

<212> DNA

<213> Rattus norvegicus

<400> 5

```
atgtcgttct tcctcagcaa cttaacaaat gactctagac tgtggaaagt aagtcataat     60
tctacggacc ttatgaattc gccagaaact ctgactcttt ctctcttttg cctgatctgt    120
ttaatgactc tcgtagctct ggtaggcagc attttctcac tggtttctct gctcaccatg    180
cagtacagaa ctgttgtgtc tatgcttgtg acctcttggt ctgtggatga tctcttgagc    240
gtcttgtcag tggccatctt catggttttg cagtggccaa gagaggcacc gggctacttc    300
cagtctctgt gtaccacctc tgccttactg tatatgtgcc agggcctctc cagcaacttg    360
aaggcgactc ttatagtctt ctacaacttt tatacgatgc acaggacagt ggtgagtcag    420
tcaagctcct ggcgatcagg acaggtactc ggtgtggctt taaccgtgtg ggcagtcagc    480
ctgctgctgg cctcgctccc gctgtgcggc tggggtgtct tcgtgcgcac gccttggggc    540
tgcctgaccg actgctccag cccatacgtg ctgctcctct tcgcggtgta cgcttccgct    600
ttcgggctcc tggcggttct ttctgtccct ctcactcacc agctgctgtg ttcagaggag    660
ccgccaagac tccatgccaa ctaccaggaa atttctcgtg gcgcctctac tcctgggacc    720
cctgctgctg ggggaagagt gctctgtctg ttgccagagg atgttgaaat ccctgctctg    780
ccaggcactg ggagctccct gagctccgac atggtgtttg caccgggtca gcctgctgcc    840
tctagcgctg gagccggcaa gcgagagaat ctctggactc cccgtggctc gagcagcttc    900
cccgtgagcc tggcacagaa gcgctttgct ttgatcctag cgctgactaa agtcatcctc    960
tggctgccta tgatgatcca catggtggta aaacacgtgg tggggtttca gagccttccg   1020
gtcgatatgc tcagctttct actaaccctg ctggccagca ctgtaacacc agtgtttgtc   1080
ttgtctaaac gatgggccca cttgccttgt ggctgcatca tcaactgcca gcctgacacc   1140
tactccgttg cattcgatgg gaaaaagtca aagaggaaag gctttgaatt caatctatcg   1200
ttccaacaaa gttatggact ctataaaatg acacatgcag attactatga cgatgatgat   1260
gaaaatccca tatcctacca caacccaaag aaatatgagt gtgaagctac aaaggagcct   1320
cgggaagaca accacggtgt gttcaatacc atcacagtag aaatcagcac cacgccaccg   1380
ctggacagcg ccacactgac cggtgtcaac aagtgcacca acacggatat ccctgagccc   1440
aaacaggccg tgagcgaaga aaagggtgcc ttttctatta aaacagaatg tgctattaac   1500
tatggagaag ccacctcgtt tgaaggtcca gagagacgac tgtcccatga ggagactcag   1560
aaaccagatc tttcagactg ggagtggtgc aggagtaagt cagaacgaac acctcgccag   1620
cgttctggtg gtgggctggc tattcccatt tgtgcgttcc aagggactgt gtctctccaa   1680
gcaccaactg ggaaaaccct ctctctttct acttatgagg tcagtgcaga agggcaaaag   1740
ataaccccgc cctctaagaa aatagaagtc tatcgatcta aaagtgttgg ccatgaaccg   1800
aactcagaag agtctccctc cacgtttgca gacaccaacg taaaaataca cttggaagtt   1860
cttgaaattt gtgacaacga tgaggcttta gacactgtgt caatcatcag caacatcagc   1920
cagtcctcta caaaagttcg gtctccttct ctgcgctact cacggaagga aaacaggttt   1980
gtctcatgcg acttaggaga aactgcctcc tactccctct tcttacccac aagtgatcct   2040
gacggagata tcaacatctc tatcccagac acagtggagg cacacagaca gaacagcagg   2100
cggcagcatc aagacaggga tggctaccag gaagagatcc agttgttaaa caaagcctac   2160
aggaaaagag aagcagaaag caagggtaac tag                                 2193
```

<210> 6

<211> 730

<212> PRT

<213> Rattus norvegicus

<400> 6

```
Met Ser Phe Phe Leu Ser Asn Leu Thr Asn Asp Ser Arg Leu Trp Lys
1               5                   10                  15
Val Ser His Asn Ser Thr Asp Leu Met Asn Ser Pro Glu Thr Leu Thr
            20                  25                  30
Leu Ser Leu Phe Cys Leu Ile Cys Leu Met Thr Leu Val Ala Leu Val
        35                  40                  45
Gly Ser Ile Phe Ser Leu Val Ser Leu Leu Thr Met Gln Tyr Arg Thr
    50                  55                  60
Val Val Ser Met Leu Val Thr Ser Trp Ser Val Asp Asp Leu Leu Ser
65                  70                  75                  80
Val Leu Ser Val Ala Ile Phe Met Val Leu Gln Trp Pro Arg Glu Ala
                85                  90                  95
Pro Gly Tyr Phe Gln Ser Leu Cys Thr Thr Ser Ala Leu Leu Tyr Met
            100                 105                 110
Cys Gln Gly Leu Ser Ser Asn Leu Lys Ala Thr Leu Ile Val Phe Tyr
        115                 120                 125
Asn Phe Tyr Thr Met His Arg Thr Val Val Ser Gln Ser Ser Ser Trp
    130                 135                 140
Arg Ser Gly Gln Val Leu Gly Val Ala Leu Thr Val Trp Ala Val Ser
145                 150                 155                 160
Leu Leu Leu Ala Ser Leu Pro Leu Cys Gly Trp Gly Val Phe Val Arg
                165                 170                 175
Thr Pro Trp Gly Cys Leu Thr Asp Cys Ser Ser Pro Tyr Val Leu Leu
            180                 185                 190
Leu Phe Ala Val Tyr Ala Ser Ala Phe Gly Leu Leu Ala Val Leu Ser
        195                 200                 205
Val Pro Leu Thr His Gln Leu Leu Cys Ser Glu Glu Pro Pro Arg Leu
    210                 215                 220
His Ala Asn Tyr Gln Glu Ile Ser Arg Gly Ala Ser Thr Pro Gly Thr
225                 230                 235                 240
Pro Ala Ala Gly Gly Arg Val Leu Cys Leu Leu Pro Glu Asp Val Glu
                245                 250                 255
Ile Pro Ala Leu Pro Gly Thr Gly Ser Ser Leu Ser Ser Asp Met Val
            260                 265                 270
Phe Ala Pro Gly Gln Pro Ala Ala Ser Ser Ala Gly Ala Gly Lys Arg
        275                 280                 285
Glu Asn Leu Trp Thr Pro Arg Gly Ser Ser Ser Phe Pro Val Ser Leu
```

290 295 300

Ala Gln Lys Arg Phe Ala Leu Ile Leu Ala Leu Thr Lys Val Ile Leu

305 310 315 320

Trp Leu Pro Met Met Ile His Met Val Val Lys His Val Val Gly Phe

325 330 335

Gln Ser Leu Pro Val Asp Met Leu Ser Phe Leu Leu Thr Leu Leu Ala

340 345 350

Ser Thr Val Thr Pro Val Phe Val Leu Ser Lys Arg Trp Ala His Leu

355 360 365

Pro Cys Gly Cys Ile Ile Asn Cys Gln Pro Asp Thr Tyr Ser Val Ala

370 375 380

Phe Asp Gly Lys Lys Ser Lys Arg Lys Gly Phe Glu Phe Asn Leu Ser

385 390 395 400

Phe Gln Gln Ser Tyr Gly Leu Tyr Lys Met Thr His Ala Asp Tyr Tyr

405 410 415

Asp Asp Asp Asp Glu Asn Pro Ile Ser Tyr His Asn Pro Lys Lys Tyr

420 425 430

Glu Cys Glu Ala Thr Lys Glu Pro Arg Glu Asp Asn His Gly Val Phe

435 440 445

Asn Thr Ile Thr Val Glu Ile Ser Thr Thr Pro Pro Leu Asp Ser Ala

450 455 460

Thr Leu Thr Gly Val Asn Lys Cys Thr Asn Thr Asp Ile Pro Glu Pro

465 470 475 480

Lys Gln Ala Val Ser Glu Glu Lys Gly Ala Phe Ser Ile Lys Thr Glu

485 490 495

Cys Ala Ile Asn Tyr Gly Glu Ala Thr Ser Phe Glu Gly Pro Glu Arg

500 505 510

Arg Leu Ser His Glu Glu Thr Gln Lys Pro Asp Leu Ser Asp Trp Glu

515 520 525

Trp Cys Arg Ser Lys Ser Glu Arg Thr Pro Arg Gln Arg Ser Gly Gly

530 535 540

Gly Leu Ala Ile Pro Ile Cys Ala Phe Gln Gly Thr Val Ser Leu Gln

545 550 555 560

Ala Pro Thr Gly Lys Thr Leu Ser Leu Ser Thr Tyr Glu Val Ser Ala

565 570 575

Glu Gly Gln Lys Ile Thr Pro Pro Ser Lys Lys Ile Glu Val Tyr Arg

580 585 590

Ser Lys Ser Val Gly His Glu Pro Asn Ser Glu Glu Ser Pro Ser Thr

595 600 605

Phe Ala Asp Thr Asn Val Lys Ile His Leu Glu Val Leu Glu Ile Cys

610 615 620

Asp Asn Asp Glu Ala Leu Asp Thr Val Ser Ile Ile Ser Asn Ile Ser

625 630 635 640

Gln Ser Ser Thr Lys Val Arg Ser Pro Ser Leu Arg Tyr Ser Arg Lys
645 650 655
Glu Asn Arg Phe Val Ser Cys Asp Leu Gly Glu Thr Ala Ser Tyr Ser
660 665 670
Leu Phe Leu Pro Thr Ser Asp Pro Asp Gly Asp Ile Asn Ile Ser Ile
675 680 685
Pro Asp Thr Val Glu Ala His Arg Gln Asn Ser Arg Arg Gln His Gln
690 695 700
Asp Arg Asp Gly Tyr Gln Glu Glu Ile Gln Leu Leu Asn Lys Ala Tyr
705 710 715 720
Arg Lys Arg Glu Ala Glu Ser Lys Gly Asn
725 730

<210> 7
<211> 248
<212> DNA
<213> Gallus gallus

<400> 7
gtatggaggc atggaaagaa caggtgattt gcattctcac ctcaccaaca taaatcaggc 60
ctgccaccat gatcagacgg gctatgctgc tgtgacagtg atggagagct gaagagaatc 120
agggtcagag actgcaggcc gctgtgatgg agatggatca ctctgtgcaa gatgttgctt 180
ttcacgtgct gctcttccaa ggttgaaaac aaagttcctc cataagattg cctcaaggct 240
ccttagta 248

<210> 8
<211> 431
<212> DNA
<213> Mus musculus

<400> 8
gtgtcttcgt gcgcacgcct tggggctgcc tgaccgactg ctccagccca tacgtgctgc 60
tcctcttcgc ggtatacgct tccgctttcg ggctcctggc ggttctctct gtccctctca 120
ctcaccagct gctgtgttcg gaggaaccgc tgagactcca tgccaactac caggaaattt 180
ctcgtggcgc ctctactcct gggacccctg ctgctgcggg aagagtgctc tgtctgccgc 240
cagaggatgt tgagatccct gctctgcgat gcactggggg atgctccccg agctccgacg 300
tggtgtttgc cccgggtcag ccggctgcct ccggcgctgg agccggcagg cgagagaatc 360
ccgggactcc ccagggcacc aacagcttcc ccttgagctt ggcacagaag cgcttttctt 420
tgattctagc g 431

<210> 9
<211> 143
<212> PRT

<213> Mus musculus

<400> 9

Val Phe Val Arg Thr Pro Trp Gly Cys Leu Thr Asp Cys Ser Ser Pro
1 5 10 15

Tyr Val Leu Leu Leu Phe Ala Val Tyr Ala Ser Ala Phe Gly Leu Leu
20 25 30

Ala Val Leu Ser Val Pro Leu Thr His Gln Leu Leu Cys Ser Glu Glu
35 40 45

Pro Leu Arg Leu His Ala Asn Tyr Gln Glu Ile Ser Arg Gly Ala Ser
50 55 60

Thr Pro Gly Thr Pro Ala Ala Ala Gly Arg Val Leu Cys Leu Pro Pro
65 70 75 80

Glu Asp Val Glu Ile Pro Ala Leu Arg Cys Thr Gly Gly Cys Ser Pro
85 90 95

Ser Ser Asp Val Val Phe Ala Pro Gly Gln Pro Ala Ala Ser Gly Ala
100 105 110

Gly Ala Gly Arg Arg Glu Asn Pro Gly Thr Pro Gln Gly Thr Asn Ser
115 120 125

Phe Pro Leu Ser Leu Ala Gln Lys Arg Phe Ser Leu Ile Leu Ala
130 135 140

<210> 10
<211> 1824
<212> DNA
<213> Homo sapiens

<400> 10

atgtctttat ttctcagtaa cttatcaaca aatgactcta gcctgtggaa agagaatcat 60
aattctacgg accttttaaa tccgccagga accctgaata tctatctttt ttgcttgaca 120
tgtctcatga cttttgcagc cttggtgggc agcatttatt cactaatttc cctgctgaaa 180
atgcagaaca gaactgttgt gtccatgctt gtggcttcct ggtctgtgga tgatctcatg 240
agcgtcctgt cggtgaccat cttcatgttt ttgcagtggc caaacgaggt ccccggttac 300
ttccaatttc tgtgcaccac ctctgcctta atgtatttat gccagggcct ctctagcaac 360
ttgaaggcga ctctcctagt ctcttacaac ttttatacga tgcacagagg tgtggggagc 420
cagacagcct ccagaagatc gggccaggtg ctcggcgtgg tgctgaccgt gtgggcagcc 480
agtctgctgc tctcggcgct cccgctgtgc ggctggggcg cctccgtgcg cacgccctgg 540
ggctgcctgg tggactgctc cagctcctac gtactattcc tctctatcgt gtacgctttg 600
gccttcggac tcctcgtggg cctctcagtc ccactcactc accgattgct gtgttcggag 660
gagccgccga gactccactc caactaccag gaaatttccc gtggagcttc aattcctggg 720
acccctccta ctgcggggag agtggtttcc ctgtccccag aggatgctcc aggcccgagt 780
ctgcggcgct ctgggggatg ctctccgagc tccgacaccg tgttcggacc gggtgcgccc 840
gctgccgctg ggctgaagc ctgcaggcgt gagaaccggg ggactctcta tggcaccagg 900

```
agcttcaccg tgagcgtagc gcagaagcgc ttcgctttga tcctagcgct tacaaaagtc    960
gtcctttggc tgcccatgat gatgcacatg gtggtccaga acgtcgtggg gtttcagagc   1020
cttcccttgg agacattcag cttcctactt accctgctgg ccaccactgt aaccccagtg   1080
tttgtcttgt ccaaacgctg gacccacttg ccctgtggct gcatcatcaa ctgcaggcag   1140
aatgcatatg cagtggcgtc cgatgggaaa aaaatcaaga gaaaaggctt tgaattcaat   1200
ctatcattcc aaaaaagtta tgggatttat aaaatagcac atgaagatta ctatgatgat   1260
gatgaaaatt ccatattcta tcacaacctg atgaactctg agtgtgaaac tacaaaagac   1320
cctcagagag acaaccgtaa catcttcaat gctataaaag tagaaatcag caccacgccc   1380
tctctggaca gctccacaca aagaggcatc aacaaatgca caaatactga tattacagaa   1440
gctaaacagg attccaacaa caaaaaggat gcgttttctg acaaaacagg aggtgatatt   1500
aactatgaag aaactacctt ttctgaaggg ccagaaagaa gactgtctca tgaagagagt   1560
cagaaaccag atctttcaga ctgggagtgg cgtaggagta aatcagaaag aacccctcgt   1620
cagcgttccg gttatgccct tgccattccc ttgtgtgcat tccaggggac tgtgtctctc   1680
catgcaccta cagggaaaac cctatctctt tctacctatg aggtaagcgc agaagggcaa   1740
aaaataactc cagcctctaa gaaaatagaa gtctatcgat ccaaaagtgt tggccatgaa   1800
ccaaactcag aagattcttc atcc                                          1824
```

```
<210> 11
<211> 608
<212> PRT
<213> Homo sapiens

<400> 11
Met Ser Leu Phe Leu Ser Asn Leu Ser Thr Asn Asp Ser Ser Leu Trp
1               5                   10                  15
Lys Glu Asn His Asn Ser Thr Asp Leu Leu Asn Pro Pro Gly Thr Leu
            20                  25                  30
Asn Ile Tyr Leu Phe Cys Leu Thr Cys Leu Met Thr Phe Ala Ala Leu
        35                  40                  45
Val Gly Ser Ile Tyr Ser Leu Ile Ser Leu Leu Lys Met Gln Asn Arg
    50                  55                  60
Thr Val Val Ser Met Leu Val Ala Ser Trp Ser Val Asp Asp Leu Met
65                  70                  75                  80
Ser Val Leu Ser Val Thr Ile Phe Met Phe Leu Gln Trp Pro Asn Glu
                85                  90                  95
Val Pro Gly Tyr Phe Gln Phe Leu Cys Thr Thr Ser Ala Leu Met Tyr
            100                 105                 110
Leu Cys Gln Gly Leu Ser Ser Asn Leu Lys Ala Thr Leu Leu Val Ser
        115                 120                 125
Tyr Asn Phe Tyr Thr Met His Arg Gly Val Gly Ser Gln Thr Ala Ser
    130                 135                 140
Arg Arg Ser Gly Gln Val Leu Gly Val Val Leu Thr Val Trp Ala Ala
145                 150                 155                 160
```

```
Ser Leu Leu Leu Ser Ala Leu Pro Leu Cys Gly Trp Gly Ala Ser Val
                165                 170                 175
Arg Thr Pro Trp Gly Cys Leu Val Asp Cys Ser Ser Ser Tyr Val Leu
            180                 185                 190
Phe Leu Ser Ile Val Tyr Ala Leu Ala Phe Gly Leu Leu Val Gly Leu
        195                 200                 205
Ser Val Pro Leu Thr His Arg Leu Leu Cys Ser Glu Glu Pro Pro Arg
    210                 215                 220
Leu His Ser Asn Tyr Gln Glu Ile Ser Arg Gly Ala Ser Ile Pro Gly
225                 230                 235                 240
Thr Pro Pro Thr Ala Gly Arg Val Val Ser Leu Ser Pro Glu Asp Ala
                245                 250                 255
Pro Gly Pro Ser Leu Arg Arg Ser Gly Gly Cys Ser Pro Ser Ser Asp
            260                 265                 270
Thr Val Phe Gly Pro Gly Ala Pro Ala Ala Ala Gly Ala Glu Ala Cys
        275                 280                 285
Arg Arg Glu Asn Arg Gly Thr Leu Tyr Gly Thr Arg Ser Phe Thr Val
    290                 295                 300
Ser Val Ala Gln Lys Arg Phe Ala Leu Ile Leu Ala Leu Thr Lys Val
305                 310                 315                 320
Val Leu Trp Leu Pro Met Met Met His Met Val Val Gln Asn Val Val
                325                 330                 335
Gly Phe Gln Ser Leu Pro Leu Glu Thr Phe Ser Phe Leu Leu Thr Leu
            340                 345                 350
Leu Ala Thr Thr Val Thr Pro Val Phe Val Leu Ser Lys Arg Trp Thr
        355                 360                 365
His Leu Pro Cys Gly Cys Ile Ile Asn Cys Arg Gln Asn Ala Tyr Ala
    370                 375                 380
Val Ala Ser Asp Gly Lys Lys Ile Lys Arg Lys Gly Phe Glu Phe Asn
385                 390                 395                 400
Leu Ser Phe Gln Lys Ser Tyr Gly Ile Tyr Lys Ile Ala His Glu Asp
                405                 410                 415
Tyr Tyr Asp Asp Asp Glu Asn Ser Ile Phe Tyr His Asn Leu Met Asn
            420                 425                 430
Ser Glu Cys Glu Thr Thr Lys Asp Pro Gln Arg Asp Asn Arg Asn Ile
        435                 440                 445
Phe Asn Ala Ile Lys Val Glu Ile Ser Thr Thr Pro Ser Leu Asp Ser
    450                 455                 460
Ser Thr Gln Arg Gly Ile Asn Lys Cys Thr Asn Thr Asp Ile Thr Glu
465                 470                 475                 480
Ala Lys Gln Asp Ser Asn Asn Lys Lys Asp Ala Phe Ser Asp Lys Thr
                485                 490                 495
Gly Gly Asp Ile Asn Tyr Glu Glu Thr Thr Phe Ser Glu Gly Pro Glu
```

500 505 510

Arg Arg Leu Ser His Glu Glu Ser Gln Lys Pro Asp Leu Ser Asp Trp

515 520 525

Glu Trp Arg Arg Ser Lys Ser Glu Arg Thr Pro Arg Gln Arg Ser Gly

530 535 540

Tyr Ala Leu Ala Ile Pro Leu Cys Ala Phe Gln Gly Thr Val Ser Leu

545 550 555 560

His Ala Pro Thr Gly Lys Thr Leu Ser Leu Ser Thr Tyr Glu Val Ser

565 570 575

Ala Glu Gly Gln Lys Ile Thr Pro Ala Ser Lys Lys Ile Glu Val Tyr

580 585 590

Arg Ser Lys Ser Val Gly His Glu Pro Asn Ser Glu Asp Ser Ser Ser

595 600 605